Paula Kotzé Gary Marsden Gitte Lindgaard
Janet Wesson Marco Winckler (Eds.)

LNCS 8120

Human-Computer Interaction – INTERACT 2013

14th IFIP TC 13 International Conference
Cape Town, South Africa, September 2013
Proceedings, Part IV

Part IV

ifip

Springer

Lecture Notes in Computer Science 8120

Paula Kotzé Gary Marsden Gitte Lindgaard
Janet Wesson Marco Winckler (Eds.)

Human-Computer Interaction – INTERACT 2013

14th IFIP TC 13 International Conference
Cape Town, South Africa, September 2-6, 2013
Proceedings, Part IV

Springer

Volume Editors

Paula Kotzé
CSIR Meraka Institute, PO Box 395, Pretoria 0001, South Africa
E-mail: pkotze1@csir.co.za

Gary Marsden
University of Cape Town, Department of Computer Science
Private Bag X3, Rondebosch 7701, Cape Town, South Africa
E-mail: gaz@cs.uct.ac.za

Gitte Lindgaard
Carleton University, 1125 Colonel By Drive, Ottawa, ON K1S 5B6, Canada
E-mail: gitte.lindgaard@carleton.ca

Janet Wesson
Nelson Mandela Metropolitan University, Department of Computing Sciences
PO Box 77000, Port Elisabeth 6031, South Africa
E-mail: janet.wesson@nmmu.ac.za

Marco Winckler
University Paul Sabatier, Institute of Research in Informatics of Toulouse (IRIT)
118 Route de Narbonne, 31062 Toulouse Cedex 9, France
E-mail: winckler@irit.fr

ISSN 0302-9743 e-ISSN 1611-3349
ISBN 978-3-642-40497-9 e-ISBN 978-3-642-40498-6
DOI 10.1007/978-3-642-40498-6
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013945558

CR Subject Classification (1998): H.5.2, H.5.3, H.3-5, I.2.10, D.2, C.2, K.3-4

LNCS Sublibrary: SL 3 – Information Systems and Application, incl. Internet/Web and HCI

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Foreword

INTERACT 2013 was the 14^{th} of a series of INTERACT international conferences supported by the International Federation for Information Processing (IFIP) Technical Committee 13 on Human–Computer Interaction.

This year, INTERACT was held in Cape Town (South Africa), organized by the Nelson Mandela Metropolitan University (Port Elizabeth) and the Meraka Institute of Council for Scientific and Industrial Research (Pretoria) in collaboration with the University of Cape Town.

The Conference theme for INTERACT 2013, "Designing for Diversity," recognizes the interdisciplinary, multidisciplinary and intercultural spirit of human–computer interaction (HCI) research and practice. The conference welcomes research and reports of practice that acknowledge diverse disciplines, abilities, cultures, and societies, and that address both the technical and social aspects of HCI. Within the broad umbrella of HCI, the conference sought contributions addressing new and emerging HCI disciplines, bridging cultural differences, and tackling important social problems.

Like its predecessors, INTERACT 2013 highlighted, to both the academic and the industrial world, the importance of the HCI discipline and its most recent breakthroughs on current applications. Both experienced HCI researchers and professionals, as well as newcomers to the HCI field, interested in designing or evaluating interactive software, developing new interaction technologies, or investigating overarching theories of HCI, found in INTERACT 2013 an exciting forum for communication with people of similar interests, to encourage collaboration and to learn.

INTERACT 2013 brought the conference to South Africa and Africa for the very first time. The African tradition of HCI focuses very much on the human and social aspects of HCI, recognizing the diversity of its people and the circumstance in which they go about their everyday lives. We hope that INTERACT 2013 will be remembered as a conference that brought the diversity of HCI research to the forefront, making the computerized world a better place for all, regardless of where they come from.

INTERACT 2013 took place 29 years after the first INTERACT held in September 1984 in London, UK. The IFIP Technical Committee 13 aims to develop the science and technology of the interaction between humans and computing devices through different Working Groups and Special Interests Groups, all of which, together with their officers, are listed within these proceedings.

We thank all the authors who chose INTERACT 2013 as the venue to publish their research. This was again an outstanding year for the conference in terms of submissions in all the technical categories, especially since the conference moved away from the traditional predominantly European venues. In total, we received 639 submissions. Of these, 270 submissions were accepted:

- 128 as full research papers
- 77 as short research papers
- 31 as interactive posters
- 2 as industrial programme papers
- 4 as panels
- 1 as a special interest group
- 1 as a tutorial
- 9 as workshops
- 9 to the African Masters Consortium
- 8 to the Doctoral Consortium

The acceptance rate for the full and short research papers was 31% and 45%, respectively.

A Programme Committee meeting consisting of the Technical Programme Chairs and the Track Chairs, as well as member of IFIP Technical Committee 13, preceded the final decision on which submissions to accept. This powerful effort was only possible thanks to the diligent work of many people. Our sincere gratitude goes to the almost 700 members of our International Programme Committee who willingly assisted and ensured the high quality of the INTERACT Conference papers was properly maintained. Although some people had to be bullied into reviewing (sorry about that), everyone submitted their reviews on time without a murmur of complaint. Thank you all for the effort that you so obviously put into this task. A special thank you must go to our Track Chairs, who put in a tremendous amount of work to ensure that quality was maintained throughout.

In addition, we have to thank the members of the Organizing Committee, the staff at the Council for Industrial and Scientific Research, Nelson Mandela Metropolitan University and the University of Cape Town for their unflagging assistance with all aspects of planning and managing the many administrative and organizational issues. We also have to thank our student volunteers for making sure that everything ran smoothly at the conference itself.

Finally, we wish to express a special thank you to the Proceedings Publication Chair, Marco Winckler, who painstakingly put this volume together.

September 2013

Paula Kotzé
Janet Wesson
(INTERACT 2013 Conference Chairs)
Gary Marsden
Gitte Lindgaard
(INTERACT 2013 Technical Programme Chairs)

IFIP TC13

Established in 1989, the International Federation for Information Processing Technical Committee on Human–Computer Interaction (IFIP TC13) is an international committee of 30 member national societies and seven Working Groups, representing specialists in human factors, ergonomics, cognitive science, computer science, design, and related disciplines. INTERACT is its flagship conference, staged biennially in different countries in the world.

IFIP TC13 aims to develop the science and technology of human–computer interaction (HCI) by encouraging empirical research, promoting the use of knowledge and methods from the human sciences in design and evaluation of computer systems; promoting better understanding of the relationship between formal design methods and system usability and acceptability; developing guidelines, models and methods by which designers may provide better human-oriented computer systems; and, cooperating with other groups, inside and outside IFIP, to promote user-orientation and humanization in system design. Thus, TC13 seeks to improve interactions between people and computers, encourage the growth of HCI research, and disseminate these benefits worldwide.

The main orientation is toward users, especially the non-computer professional users, and how to improve human–computer relations. Areas of study include: the problems people have with computers; the impact on people in individual and organizational contexts; the determinants of utility, usability, and acceptability; the appropriate allocation of tasks between computers and users; modeling the user to aid better system design; and harmonizing the computer to user characteristics and needs.

While the scope is thus set wide, with a tendency toward general principles rather than particular systems, it is recognized that progress will only be achieved through both general studies to advance theoretical understanding and specific studies on practical issues (e.g., interface design standards, software system consistency, documentation, appropriateness of alternative communication media, human factors guidelines for dialogue design, the problems of integrating multimedia systems to match system needs and organizational practices, etc.).

In 1999, TC13 initiated a special IFIP Award, the Brian Shackel Award, for the most outstanding contribution in the form of a refereed paper submitted to and delivered at each INTERACT. The award draws attention to the need for a comprehensive human-centered approach in the design and use of information technology in which the human and social implications have been taken into account. Since the process to decide the award takes place after papers are submitted for publication, the award is not identified in the proceedings.

IFIP TC13 stimulates working events and activities through its Working Groups (WGs) and Special Interest Groups (SIGs). WGs and SIGs consist of HCI experts from many countries, who seek to expand knowledge and find solutions to HCI issues and concerns within their domains, as outlined below.

- WG13.1 (Education in HCI and HCI Curricula) aims to improve HCI education at all levels of higher education, coordinate and unite efforts to develop HCI curricula and promote HCI teaching.
- WG13.2 (Methodology for User-Centered System Design) aims to foster research, dissemination of information and good practice in the methodical application of HCI to software engineering.
- WG13.3 (HCI and Disability) aims to make HCI designers aware of the needs of people with disabilities and encourage development of information systems and tools permitting adaptation of interfaces to specific users.
- WG13.4 (also WG2.7) (User Interface Engineering) investigates the nature, concepts and construction of user interfaces for software systems, using a framework for reasoning about interactive systems and an engineering model for developing user interfaces.
- WG13.5 (Human Error, Safety and System Development) seeks a framework for studying human factors relating to systems failure, develops leading-edge techniques in hazard analysis and safety engineering of computer-based systems, and guides international accreditation activities for safety-critical systems.
- WG13.6 (Human-Work Interaction Design) aims at establishing relationships between extensive empirical work-domain studies and HCI design. It will promote the use of knowledge, concepts, methods and techniques that enable user studies to procure a better apprehension of the complex interplay between individual, social and organizational contexts and thereby a better understanding of how and why people work in the ways that they do.
- WG13.7 (Human–Computer Interaction and Visualization) is the newest of the working groups under the TC13. It aims to establish a study and research program that will combine both scientific work and practical applications in the fields of HCI and visualization. It will integrate several additional aspects of further research areas, such as scientific visualization, data mining, information design, computer graphics, cognition sciences, perception theory, or psychology, into this approach.
- SIG 13.1 (HCI and International Development) aims to promote the application of interaction design research, practice and education to address the needs, desires and aspirations of people in the developing world; support and develop the research, practice and education capabilities of HCI institutions and organizations based in the developing world; develop links between the HCI community in general, and IFIP TC13 in particular, with other relevant communities involved in development, especially IFIP WG 9.4 Computers in Developing Countries.

- SIG 13.2 (Interaction Design and Children) aims to provide a forum for all things relating to interaction design and HCI where the intended users or appropriators of the technology or service are children. The definition of children is broad rather than narrow, including toddlers and teenagers, but the core work, currently at least, is with children in junior schools.

New Working Groups and Special Interest Groups are formed as areas of significance to HCI arise. Further information is available at the IFIP TC13 website: http://www.tc13.org

IFIP TC13 Members

Australia
Judy Hammond
Australian Computer Society

Australia
Andreas Holzinger
Austrian Computer Society

Belgium
Monique Noirhomme-Fraiture
Fédération des Associations
Informatiques de Belgique

Brazil
Simone Diniz Junqueira Barbosa
Brazilian Computer Society (SBC)

Bulgaria
Kamelia Stefanova
Bulgarian Academy of Sciences

Canada
Heather O'Brien
Canadian Information Processing
Society

Chile
Jaime Sánchez
Chilean Society of Computer Science

China
Chengqing Zong
Chinese Institute of Electronics

Cyprus
Panayiotis Zaphiris
Cyprus Computer Society

Czech Republic
Vaclav Matousek
Czech Society for Cybernetics
and Informatics

Denmark
Torkil Clemmensen
Danish Federation
for Information Processing

Finland
Kari-Jouko Räihä
Finnish Information Processing
Association

France
Philippe Palanque
Société des Electriciens
et des Electroniciens (SEE)

Germany
Tom Gross
Gesellschaft für Informatik

Hungary
Cecilia Sik Lanyi
John V. Neumann Computer Society

Iceland
Marta Kristin Larusdottir
The Icelandic Society
for Information Processing (ISIP)

India
Anirudha Joshi
Computer Society of India

Ireland
Liam J. Bannon
Irish Computer Society

Italy
Fabio Paternò
Italian Computer Society

Japan
Yoshifumi Kitamura
Information Processing Society of Japan

Kenya
Dan Orwa
Computer Society of Kenya

Korea
Gerry Kim
KIISE

Malaysia
Chui Yin Wong
Malaysian National Computer Confederation

New Zealand
Mark Apperley
New Zealand Computer Society (NZCS)

Nigeria
Chris C. Nwannenna
Nigeria Computer Society

Norway
Dag Svanes
Norwegian Computer Society

Poland
J.L. Kulikowski
Poland Academy of Sciences

Portugal
Joaquim A. Jorge
Associacão Portuguesa para o Desenvolvimento da Sociedade da Informação (APDSI)

Singapore
Henry B.L. Duh
Singapore Computer Society

South Africa
Janet L. Wesson
The Computer Society of South Africa

Spain
Julio Abascal
Asociación de Técnicos de Informática (ATI)

Sweden
Jan Gulliksen
Swedish Computer Society

Switzerland
Solange Ghernaouti
Swiss Federation for Information Processing

The Netherlands
Gerrit C. van der Veer
Nederlands Genootschap voor Informatica

UK
Andy Dearden
British Computer Society

USA-Based
John Karat
Association for Computing Machinery (ACM)

Expert Members
Paula Kotzé (South Africa)
Gitte Lindgaard (Australia)
Marco Winckler (France)
Masaaki Kurosu (Japan)
Nikos Avouris (Greece)
Zhengjie Liu (China)

Working Group Chairpersons

WG13.1 (Education in HCI and HCI Curricula)
Lars Oestreicher, Sweden

WG13.2 (Methodology for User-Centered System Design)
Peter Forbrig, Germany

WG13.3 (HCI and Disability)
Gerhard Weber, Germany

WG13.4 (also 2.7) (User Interface Engineering)
Fabio Paternò, Italy

WG13.5 (Human Error, Safety, and System Development)
Chris Johnson, UK

WG13.6 (Human-Work Interaction Design)
Torkil Clemmensen, Denmark

WG13.7 (Human Computer Interaction and Visualization)
Achim Ebert, Germany

SIG 13.1 – HCI and International Development
Andy Dearden, UK

SIG 13.2 – Interaction Design and Children
Janet Read, UK

Conference Organizing Committee

INTERACT 2013
Conference Chairs
Paula Kotzé, South Africa
Janet Wesson, South Africa

Technical Programme Chairs
Gary Marsden, South Africa
Gitte Lindgaard, Australia

Long Research Papers Chairs
Gary Marsden, South Africa
Gitte Lindgaard, Australia

Short Research Papers and Interactive Posters Chairs
Peter Forbrig, Germany
Monique Noirhomme, Belgium

Industrial Programme Chairs
Susan Dray, USA
Shikoh Gitau, Kenya
Helga Letowt-Vorbek, South Africa

Keynotes Chairs
Gary Marsden, South Africa
Paula Kotzé, South Africa
Janet Wesson, South Africa

Workshops and Tutorials Chair
Helene Gelderblom, South Africa

Doctoral Consortium Chairs
Philippe Palanque, France
Darelle van Greunen, South Africa
Paula Kotzé, South Africa

African Masters Consortium Chairs
Darelle van Greunen, South Africa
Paula Kotzé, South Africa

Panels and Special Interest Groups Chairs
Andy Dearden, UK
Hannah Thinyane, South Africa
Judy van Biljon, South Africa

Demonstrations and Interactive Experiences Chairs
Stacy Hobson, USA
Kasper Jensen, Namibia

Proceedings Publication Chair
Marco Winckler, France

Student Volunteer Chairs
Thomas Reitmaier, Germany
Funmi Adebesin, South Africa
Darelle van Greunen, South Africa

General Conference Administration and Logistics
Marlene Jivan, South Africa
Alida Veldsman, South Africa

International Programme Committee Members

Julio Abascal, Spain
Al Mahmud Abdullah, The Netherlands
Silvia Abrahão, Spain
Funmi Adebesin, South Africa
Sisira Adikari, Australia
Ana Paula Afonso, Portugal
Jose Luis Alba, Spain
Jason Alexander, UK
Dzmitry Aliakseyeu, The Netherlands
Majed Alshamari, Saudi Arabia
Margarita Anastassova, France
Craig Anslow, New Zealand
Jean-Yves Antoine, France
Mark Apperley, New Zealand
Caroline Appert, France
Nathalie Aquino, France
Liliana Ardissono, Italy
Carmelo Ardito, Italy
Pedro Arezes, Portugal
Myriam Arrue Recondo, Spain
Mark Ashdown, UK
Nikolaos Avouris, Greece
Chris Baber, UK
Cedric Bach, France
Jonathan Back, UK
Sebastiano Bagnara, Italy
Gilles Bailly, Germany
Martina Balestra, USA
Emilia Barakova, The Netherlands
Simone Barbosa, Brazil
Shaowen Bardzell, USA
Javier Bargas-Avila, Switzerland
Louise Barkhuus, Sweden
Barbara Rita Barricelli, Italy
Gil Barros, Brazil
Remi Bastide, France
Eric Baumer, USA
Gordon Baxter, UK
Russell Beale, UK
Elke Beck, Austria
Yacine Bellik, France
David Benyon, UK
Francois Berard, France
Regina Bernhaupt, France
Olav W. Bertelsen, Denmark
Nigel Bevan, UK
Ganesh Bhutkar, India
Nicola Bidwell, South Africa
Matthew Bietz, USA
Mark Billinghurst, UK
Dorrit Billman, USA
Thomas Binder, Denmark
Jeremy Birnholtz, USA
Fernando Birra, Portugal
Renaud Blanch, France
Marco Blumendorf, Germany
Mads Bødker Bødker, Denmark
Cristian Bogdan, Sweden
Davide Bolchini, USA
Birgit Bomsdorf, Germany
Rodrigo Bonacin, Brazil
Marcela Borge, USA
Sebastian Boring, Denmark
Aviaja Borup, Denmark
Paolo Bottoni, Italy
Matt-Mouley Bouamrane, UK
Don Bouwhuis, The Netherlands
Chris Bowers, UK
Doug Bowman, USA
Giorgio Brajnik, Italy
Willem-Paul Brinkman, The Netherlands
Gregor Broll, Germany
Judith Brown, Canada
Nick Bryan-Kinns, UK
Sabin-Corneliu Buraga, Romania
Marina Buzzi, France
Maria Claudia Buzzi, Italy
Daragh Byrne, USA

Cristina, Cachero, Spain
Paul Cairns, UK
Jeff Calcaterra, USA
Karen Calteaux, South Africa
Licia Calvi, The Netherlands
Eduardo Calvillo Gamez, Mexico
José Campos, Portugal
Pedro Campos, Portugal
Maria-Dolores Cano, Spain
Cinzia Cappiello, Italy
Robert Capra, USA
Jorge Cardoso, Portugal
Stefan Carmien, Spain
Maria Beatriz Carmo, Portugal
Luis Carriço, Portugal
Tiziana Catarci, Italy
Augusto Celentano, Italy
Teresa Chambel, Portugal
Olivier Chapuis, France
Monchu Chen, Portugal
Yunan Chen, USA
Kelvin Cheng, Singapore
Keith Cheverst, UK
Sonia Chiasson, Canada
Yoram Chisik, Portugal
Luca Chittaro, Italy
Yu-kwong Chiu, USA
Georgios Christou, Cyprus
Andrea Civan Hartzler, USA
Torkil Clemmensen, Denmark
Todd Cochrane, New Zealand
Luis Coelho, Portugal
Nathalie Colineau, Australia
Karin Coninx, Belgium
Stéphane Conversy, France
Michael Cooke, Ireland
Nuno Correia, Portugal
Maria Francesca Costabile, Italy
Celine Coutrix, France
Nadine Couture, France
Anna Cox, UK
Leonardo Cunha de Miranda, Brazil
Edward Cutrell, India
Raimund Dachselt, Germany
Carla Maria Dal Sasso Freitas, Brazil
Peter Dalsgaard, Germany
Jose Danado, Portugal
Chi Tai Dang, Germany
Peter Dannenmann, Germany
John Darzentas, Greece
Berardina De Carolis, Italy
Tjerk de Greef, The Netherlands
Elena de la Guia Cantero, Spain
Alexander De Luca, Germany
Giorgio De Michelis, Italy
Luigi De Russis, Italy
Boris de Ruyter, The Netherlands
Marco de Sá, USA
Clarisse de Souza, Brazil
Ruth de Villiers, South Africa
Andy Dearden, UK
Alexandre Demeure, France
Parag Deshpande, Sweden
Heather Desurvire, USA
Nicoletta Di Blas, Italy
Ines Di Loreto, Italy
Paulo Dias, Portugal
Lily Diaz, Finland
Claire Diederich, Belgium
Anke Dittmar, Germany
Tanja Döring, Germany
Susan Dray, USA
Carlos Duarte, Portugal
Emmanuel Dubois, France
Cathy Dudek, Canada
Andreas Duenser, New Zealand
Sophie Dupuy-Chessa, France
Hamid Ekbia, USA
Linda Elliott, USA
Niklas Elmqvist, USA
Thomas, Erickson, USA
Dominik Ertl, Austria
Parisa Eslambolchilar, UK
Giorgio Faconti, Italy
João Falcão e Cunha, Portugal
Pedro Faria Lopes, Portugal
Robert Farrell, USA
Ava Fatah gen Schieck, UK
Jean-Daniel Fekete, France
Daniel Felix, Switzerland

Xavier Ferre, Spain
Mirko Fetter, Germany
Sebastian Feuerstack, Brazil
Danyel Fisher, USA
Francisco Florez-Revuelta, UK
Daniela Fogli, Italy
Eelke Folmer, USA
Joan Fons, Spain
Manuel Joao Fonseca, Portugal
Peter Forbrig, Germany
Derek Foster, UK
Marcus Foth, Australia
Teresa Franqueira, Portugal
Andre Freire, Brazil
Mathias Frisch, Germany
Dominic Furniss, UK
Luigi Gallo, Italy
Teresa Galvao, Portugal
Nestor Garay-Vitoria, Spain
Roberto García, Spain
Vaibhav Garg, USA
Anant Bhaskar Garg, USA
Jose Luis Garrido, Spain
Franca Garzotto, Italy
Jean-marc Gauthier, USA
Miguel Gea, Spain
Patrick Gebhard, Germany
Sven Gehring, Germany
Helene Gelderblom, South Africa
Cristina Gena, Italy
Nahum Gershon, USA
Werner Geyer, USA
Florian Geyer, Germany
Gautam Ghosh, Norway
Patrick Girard, France
Shikoh Gitau, Kenya
Vlado Glavinic, Croatia
Kentaro Go, Japan
Daniel Gonçalves, Portugal
Jose Luis Gonzalez Sanchez, Spain
Alice Good, UK
Phil Gosset, USA
Nitesh Goyal, USA
Cornelia Graf, Austria
Nicholas Graham, Canada
Andrina Granic, Croatia
Raphael Grasset, Austria
Olivier Grisvard, France
Tom Gross, Germany
Joshua Gross, USA
John Grundy, Australia
Tiago Guerreiro, Portugal
Yves Guiard, France
Nuno Guimaraes, Portugal
Jan Gulliksen, Sweden
Sainath Gupta, India
Adriana Mihaela Guran, Romania
Francisco Luis Gutierrez Vela, Spain
Mieke Haesen, Belgium
Hans Hagen, Germany
Lars Hallnäs, Sweden
Kim Halskov, Denmark
Martin Halvey, UK
Judy Hammond, Australia
Mark Hancock, UK
Richard Harper, UK
Joseph Hategekimana, Norway
Kirstie Hawkey, Canada
Elaine Hayashi, Brazil
Ruediger Heimgaertner, Germany
Ingi Helgason, UK
Bart Hengeveld, The Netherlands
Michael Herczeg, Germany
Marlien Herselman, South Africa
Morten Hertzum, Denmark
Wilko Heuten, Germany
Michael Hildebrandt, Norway
Martin Hitz, Austria
Stacy Hobson, USA
Christina Hochleitner, Austria
Stefan Holmlid, Sweden
Clemens Holzmann, Austria
Kasper Hornbæk, Denmark
Heiko Hornung, Brazil
John Hosking, Australia
Juan Pablo Hourcade, USA
Ko-Hsun Huang, Taiwan
William Hudson, UK
Stéphane Huot, France
Tim Hussein, Germany

Ebba Hvannberg, Iceland
Ioanna Iacovides, UK
Junko Ichino, Japan
Netta Iivari, Finland
Shamsi Iqbal, USA
Petra Isenberg, France
Minna Isomursu, Finland
Ido A. Iurgel, Germany
Ole Sejer Iversen, Denmark
Robert Jacob, USA
Mikkel Jakobsen, Denmark
Jacek Jankowski, Ireland
Anders Jansson, Sweden
Kasper Jensen, Namibia
Chris Johnson, UK
Simon Jones, UK
Joaquim Jorge, Portugal
Rui Jose, Portugal
Martino Jose Mario, Brazil
Anirudha Joshi, India
Christophe Jouffrais, France
Marko Jurmu, Japan
Don Kalar, USA
Martin Kaltenbrunner, Austria
Santosh Kalwar, Finland
Matthew Kam, USA
Bridget Kane, Ireland
Victor Kaptelinin, Sweden
Mayur Karnik, Portugal
Kostas Karpouzis, Greece
Konstantinos Kazakos, Australia
Ryan Kelly, UK
Pramod Khambete, India
Vassilis-Javed Khan, The Netherlands
Olga Khroustaleva, USA
Hyungsin Kim, USA
Jayne Klenner, UK
Gerhard Knolmayer, Switzerland
Christophe Kolski, France
Ilpo Koskinen, Australia
Paula Kotzé, South Africa
Kostas Koukouletsos, Greece
Subramanian Krishnamurthy, India
Antonio Krüger, Germany
Sari Kujala, Finland
Hannu Kukka, Finland
Juliusz Kulikowski, Poland
Ann Lantz, Sweden
Rosa Lanzilotti, Italy
Astrid Twenebowa Larssen, Ghana
Marta Larusdottir, Iceland
Yann Laurillau, France
Marco Lazzari, Italy
Eric Lecolinet, France
Jong-Seok Lee, Korea
Juha Leino, Finland
Teemu Heikki Anttoni Leinonen, Finland
Jair Leite, Brazil
Barbara, Leporini, Italy
Sophie Lepreux, France
Helga Letowt-Vorbek, South Africa
Stefano Levialdi, Italy
Henry Lieberman, USA
Gitte Lindgaard, Australia
Zhengjie Liu, China
Sara Ljungblad, Sweden
Steffen Lohmann, Germany
Nicole Lompré, France
Víctor M. López Jaquero, Spain
Fabien Lotte, France
Stephanie Ludi, USA
Bernd Ludwig, Germany
Andreas Luedtke, Germany
Christopher Lueg, Australia
Artur Lugmayr, USA
Jo Lumsden, UK
Christof Lutteroth, New Zealand
Kris Luyten, Belgium
Saturnino Luz, Ireland
Baillie Lynne, Austria
José A. Macías, Spain
Allan MacLean, UK
Rui Madeira, Portugal
Joaquim Madeira, Portugal
Christian Maertin, Germany
Charlotte Magnusson, Sweden
Ana Gabriela Maguitman, Argentina
Murni Mahmud, Malaysia
Angela Mahr, Germany

Carleen Mailtland, USA
Alessio Malizia, UK
Jani Mäntyjärvi, Finland
Javier Marco, Spain
Ana-Maria Marhan, Romania
Gary Marsden, South Africa
Mark Marshall, UK
Hannah Marston, Germany
Benoît Martin, France
Diego Martinez, Spain
Célia Martinie, France
Masood Masoodian, New Zealand
Sara Mastro, USA
Maristella Matera, Italy
Akhil Mathur, India
Max-Emanuel Maurer, Germany
Eva Mayr, Austria
Davide Mazza, Italy
Emanuela Mazzone, UK
Scott McCrickard, USA
Tom McEwan, UK
Kevin McGee, Singapore
Indrani Medhi, USA
Gerrit Meixner, Germany
Guy Melancon, France
Alex Mitchell, Singapore
Jose Pascual Molina Masso, Spain
Tonja Molin-Juustila, Finland
Francisco Montero Simarro, Spain
Pedro Moreira, Portugal
Ann Morrison, Denmark
Christiane Moser, Austria
Omar Mubin, Australia
Florian Mueller, USA
Maurice Mulvenna, UK
Dianne Murray, UK
Rod Murray-Smith, UK
Moyen Mustaquim, Sweden
Linda Napoletano, Italy
David Navarre, France
Luciana Nedel, Brazil
Vania Neris, Brazil
Jeff Nichols, USA
Colette Nicolle, UK
Laurence Nigay, France
Erik G. Nilsson, Norway
Valentina Nisi, Portugal
Leonel Nobrega, Portugal
Sylvie Noel, Italy
Manuel Noguera, Portugal
Monique Noirhomme, Belgium
Nuno Nunes, Portugal
Clemens Nylandsted Klokmose, Denmark
Michael O'Grady, Ireland
Kenton O'Hara, USA
Marianna Obrist, UK
Lars Oestreicher, Sweden
Amy Ogan, USA
Marta Oliva, Spain
Eva Oliveira Gonçalves, Portugal
Stefan Oppl, Austria
Lidia Oshlyansky, Sweden
Nuno Otero, Sweden
Benoit Otjacques, Luxembourg
Saila Ovaska, Finland
Ana Paiva, Portugal
Philippe Palanque, France
Jose Ignacio Panach Navarrete, Spain
Emanuele Panizzi, Italy
Nadia Pantidi, UK
Luca Paolino, Italy
Eleftherios Papachristos, Greece
Narcis Pares, Spain
Tessy Pargman, Sweden
Timo Partala, Finland
Avi Parush, Canada
Oscar Pastor, Spain
Fabio Paternò, Italy
Andrew Patrick, Canada
Celeste Lyn Paul, USA
Andriy Pavlovych, Canada
Annelise Mark Pejtersen, Denmark
Helen Louise Petrie, UK
Greg Phillips, Canada
Lara Piccolo, Brazil
Emmanuel Pietriga, Chile
Thomas Pietrzak, France
Marcelo Pimenta, Brazil
Ian Pitt, Ireland

Fabio Pittarello, Italy
Andreas Pleuß, Ireland
Klaus Pohl, Germany
Roberto Polillo, Italy
Vesna Popovic, Australia
Dorin Mircea Popovici, Romania
Roman Popp, Austria
Benjamin Poppinga, Germany
Chris Power, UK
Mathura Prasad Thapliyal, India
Raquel Prates, Brazil
Costin Pribeanu, Romania
Angel Puerta, USA
Andreas Pusch, Germany
Alexandra Queirós, Portugal
Victor M.R. Penichet, Spain
Kari-Jouko Räihä, Finland
Nitendra Rajput, India
Ismo Rakkolainen, Finland
David Raneburger, Austria
Roberto Ranon, Italy
Dimitrios Raptis, Denmark
Matthias Rauterberg, The Netherlands
Janet Read, UK
Madhu Reddy, USA
Stuart Reeves, UK
Patrick Reignier, France
Rene Reiners, Germany
Thomas Reitmaier, Germany
Christian Remy, Switzerland
Karen Renaud, UK
Malte Ressin, UK
Simos Retalis, Greece
Yann Riche, USA
Thomas Rist, Germany
Guillaume Rivière, France
Toni Robertson, Australia
Simon Robinson, UK
Carsten Röcker, Germany
Armanda Rodrigues, Portugal
Yvonne Rogers, UK
Markus Rohde, Germany
Teresa Romão, Portugal
Mario Romero, Sweden
Jose, Rouillard, France
Mark Rouncefield, UK
Nicolas Roussel, France
Banani Roy, Canada
Jaime Ruiz, USA
Saqib Saeed, Pakistan
Pascal Salembier, France
Antti Salovaara, Finland
Krystian Samp, Ireland
Paulo Sampaio, Brazil
Jaime Sanchez, Chile
Bengt Sandblad, Sweden
Frode Eika Sandnes, Norway
Vagner Santana, Brazil
Carmen Santoro, Italy
Teresa Sarmento, Portugal
Corina Sas, UK
Christine Satchell, Australia
Cheryl Savery, Canada
Anthony Savidis, Greece
Dominique L. Scapin, France
Thomas Schlegel, Germany
Michael Schmidt, Germany
Kevin Schneider, Canada
Johannes Schöning, Belgium
Gerhard Schwabe, Switzerland
Marc Seissler, Germany
Dev Sen, Canada
Montserrat Sendin Veloso, Spain
Tacha Serif, Turkey
Ehud Sharlin, Canada
Aditi Sharma Grover, South Africa
Moushumi Sharmin, USA
Huihui Shi, Germany
Cecília Sik Lányi, Hungary
Bruno Silva, Brazil
Paula Alexandra Silva, Portugal
Frutuoso Silva, Portugal
Joao Carlos Silva, Portugal
Milene Silveira, Brazil
Gavin Robert Sim, UK
Carla Simone, Italy
Mikael B. Skov, Denmark
Andrew Smith, South Africa
Brian Smith, USA
David Smith, USA

Danny Soroker, USA
Kenia Sousa, Belgium
Martin Spindler, Germany
Jan Stage, Denmark
Christian Stary, Austria
John Stasko, USA
Markus Stolze, Switzerland
Simone Stumpf, UK
Noi Sukaviriya, USA
S. Shyam Sundar, USA
Alistair Sutcliffe, UK
David Swallow, UK
Anthony Tang, Canada
Charlotte Tang, USA
John Tang, USA
Jean-Claude Tarby France
Franck Tarpin-Bernard, France
Deborah Tatar, USA
Luis Teixeira, Portugal
Carlos Teixeira, Portugal
António Teixeira, Puerto Rico
Harold Thimbleby, UK
Hannah Thinyane, South Africa
Jakob Tholander, Sweden
Martin Tomitsch, Australia
Ilaria Torre, Italy
Stefan Trausan-Matu, Romania
Manfred Tscheligi, Austria
Nikolaos Tselios, Greece
Simon Tucker, UK
Susan Turner, UK
Phil Turner, UK
Leon Urbas, Germany
Kaisa Väänänen-Vainio-Mattila, Finland
Heli Väätäjä, Finland
Teija Vainio, Finland
Leonel Valbom, Portugal
Pedro Valderas, Spain
Judy van Biljon, South Africa
Thea van der Geest, The Netherlands
Darelle van Greunen, South Africa
Bradley van Tonder, South Africa
Lexvan Velsen, The Netherlands
Lynette van Zijl, South Africa
Jean Vanderdonckt, Belgium
Radu-Daniel Vatavu, Romania
Manuel Veit, France
Ángel Velázquez-Iturbide, Spain
Janneke Verhaegh, The Netherlands
Arnold P.O.S. Vermeeren, The Netherlands
Bart Vermeersch, Belgium
Jo Vermeulen, Belgium
Nadine Vigouroux, France
Giuliana Vitiello, Italy
Arnd Vitzthum, Germany
Stephen Voida, USA
Jörg Voskamp, Germany
Ljiljana Vukelja, Switzerland
Dhaval Vyas, India
QuanWei Wang, China
Ian Warren, New Zealand
Gerhard Weber, Germany
Gerald Weber, New Zealand
Carolyn Wei, USA
Alexandra Weilenmann, Sweden
Astrid Weiss, Austria
Janet Wesson, South Africa
Mikael Wiberg, Sweden
Peter Wild, UK
David Wilfinger, Austria
Marco Winckler, France
Heike Winschiers-Theophilus, Namibia
Wolfgang Woerndl, Germany
Peter Wolkerstorfer, Austria
Michael Wright, UK
Min Wu, USA
Heng Xu, USA
Yeliz Yesilada, UK
James Young, Canada
Ray Yun, USA
Massimo Zancanaro, Italy
Xiaolong (Luke) Zhang, USA
Martina Ziefle, Germany
Jürgen Ziegler, Germany
Gottfried Zimmermann, Germany
Martin Zimmermann, Germany

Additional Reviewers

Álvaro Fernández
Anke Lehmann
Asbjørn Følstad
Burkay Sucu
Carlos Pereira
Cecilie Brynskov
Diana Lemme
Florian Lettner
Haiyan Jia
Ilaria Liccardi
J.L. Hiah
Jiaqi Nie
Jo Iacovides
Luciana Cardoso de Castro Salgado
Luis Duarte
Markel Vigo
Myriam Lewkowicz
Rui Madeira
Rui Nóbrega
Samuel Silva
Sandy Gould
Sofia Reis
Thomas Birn
Timm Linder
Vinícius Segura
Xavier Le Pallec
Yasmin Al Zokari

Table of Contents – Part IV

Long and Short Papers (Continued)

Supporting Physical Activity

Supporting Shared Activities

Sustainability

Tabletop Computing

Text Comprehensibility

Tracking Eyes and Head

Usability Evaluation and Technology Acceptance

User Preferences and Behaviour

User Requirements Capture and Analysis

UX in Work/Educational Contexts

Voice/Sound-Based Computing

Interactive Posters

Industry Tracks

Panels

Special Interests Groups

Tutorial

Workshops

The Development and Evaluation of an Interactive System for Age Related Musculoskeletal Rehabilitation in the Home

Mobolaji Ayoade, Stephen Uzor, and Lynne Baillie

Interactive and Trustworthy Technologies Group, School of Engineering and Built Environment, Glasgow Caledonian University,
70 Cowcaddens Road, UK
{Mobolaji.Ayoade,Stephen.Uzor,L.Baillie}@gcu.ac.uk

Abstract. This paper describes a series of user studies carried out to investigate the usability, significance, and acceptance of two visualization tools designed to improve the quality of, and adherence to home-based exercise programmes for musculoskeletal rehabilitation. The core functionality of these visualization tools enabled the users to observe the optimal way to perform their exercises via a mannequin, and receive feedback on their own movements through the use of body worn sensors. Before full deployment in the home, two user studies were carried out in the laboratory, and then two in the home with seniors who had recently undergone musculoskeletal rehabilitation using a standard care paper based booklet in the home. Our key findings suggest that by using the visualization tools the participants were able to overcome the major limitations of standard care; and that these tools were considered by the users to be useful in encouraging participation in home exercise.

Keywords: Home rehabilitation, inertial motion sensors, older adults, visualizations and musculoskeletal conditions.

1 Introduction

Musculoskeletal conditions are the most common cause of physical disability and severe long-term pain in older adults [1]. In particular, the degeneration of muscles and joints due to age can increase a person's risk of falling; and possibly lead to osteoarthritis (excessive joint pain) in the knees or hips – with a replacement of these joints being the recommended solution. The two conditions we focus on in this paper are rehabilitation after knee surgery or after a fall. These conditions normally affect older adults or seniors. For example, in Scotland, the average age for adults undergoing knee replacement surgery is 68 years [2] and over 30% of adults over the age of 65 suffer a fall at least once a year. These impairments in muscle strength and balance can have serious repercussions with regards to the independence and confidence of older adults [3].

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 1–18, 2013.

Exercise programmes are often used – both as preventive, and rehabilitative means – to help elderly patients return to functioning physical levels [4]. In hospitals and community-based rehabilitation centers, strategies involving the use of exercise for rehabilitation have been very successful [5-6]. However, in the home setting, recovery of balance and strength can be challenging, as individuals often do not adhere to rehabilitation programmes thereby making this type of intervention less effective [7-8]. In addition, the individuals sometimes do not correctly interpret the instructions and health care professionals are not aware of this lack of understanding until the patient returns to the hospital or rehabilitation clinic for a follow-up appointment.

The consequence of this lack of adherence and poor quality performance has serious implications for the health and independence of the individual. In some cases, seniors who have had a recent injurious fall may feel isolated and fearful of future falls, thereby limiting their movement and thereby actually increasing their fall risk [9-10]. In another example, people who have undergone knee replacement surgery may not recover full range of movement in the new joint(s) thereby limiting their knee function and having a negative impact on their quality of lives - if they do not exercise the joint(s) regularly as recommended by health experts [11]. In order to address these issues, it is important to identify what the short comings of standard home rehabilitation are, and investigate ways in which greater participation in home exercise can be encouraged.

One of the ways through which rehabilitation can be enhanced, we believe, is through the use of visual feedback or visualization. By visual feedback, we mean the capture of the user's movement, then presenting it in a form that is easily understandable and that indicates how well the user's movement matches the ideal movement. In previous research studies, visual feedback has been used to provide biomechanical feedback on user's performance during gait retraining [12], and the use of visual feedback to retrain balance in the elderly after a stroke [13]. However to facilitate this kind of feedback, specialized equipment and staff are often required to capture the patients' movements. Numerous studies investigating visual rehabilitation have therefore been limited to laboratory or clinical settings due to these factors [14]. To investigate the potential of visual feedback to aid rehabilitation in the home setting, there is a need to design a system capable of providing the user with meaningful feedback, while being easy to use and manage by seniors.

This paper describes a series of usability studies carried out to evaluate two distinct visualization tools to improve musculoskeletal rehabilitation for seniors using relatively cheap wireless motion capture devices. Both of these tools were designed to provide de-personalized visualizations of user movements through the use of a digital mannequin. Through the use of these systems, it is possible for the users to see the ideal range and pace of motion necessary for effective recovery, and see a visualization of their own movements during home exercise. We investigate through this research, the usability and acceptance of the visualization tools in two musculoskeletal conditions with different requirements – falls and knee replacement rehabilitation.

2 Related Work

Uzor et al. [15] identified one of the barriers to the effective use of current home rehabilitation tools was the inability to maintain motivation throughout the rehabilitation programme, which usually extends over long periods, due to the passive nature of current tools such as exercise booklets and DVDs. In addition, the absence of a physiotherapist and recording of exercises performed has been reported as other factors affecting adherence to home exercise programmes [16]. As a result, there has been a growing interest in the research community to seek technological solutions to these limitations [14], [17], [18]. According to Zhou [14], however, technologies for rehabilitation have not been accepted and adopted in the home in the past, especially by seniors for the following reasons: the systems have required professional help to set them up, occupy large or fixed spaces, are quite expensive and have poor human computer interaction.

Recently, there has been a growing interest among researchers and health professionals in using commercial systems such as the Nintendo Wii as rehabilitation tools for various conditions including musculoskeletal conditions [19]. Video games also have the potential of improving motivation and adherence to rehabilitation programs by facilitating the enjoyability of physical activities but they do not necessarily promote correct performance of rehabilitation exercises as the game controllers capture users' gestures or gross movement patterns [15], [19], [20]. Consequently, researchers and game enthusiasts are currently making attempts to develop games specifically designed for rehabilitation e.g. SilverPromenade [21] and Animal feeder [22]. And although there is evidence of increasing acceptance of commercial video game technologies by seniors for rehabilitation [23], [24], as yet the usability of such systems in an autonomous home environment is yet to be reported - all the studies have focused on rehabilitation centers or outpatient hospital departments and require help to setup. Ziefle & Röcker [25] believed that by understanding the capabilities, limitations, needs and preferences of seniors; researchers and designers could produce acceptable technologies for home rehabilitation.

Theng et al. [23] identified ease of use as one of the factors to be considered when designing technology for seniors. Although, depth cameras such as the Microsoft Kinect have the potential to be easy to use by older adults [22], camera based systems face the problem of occlusion and need they require at least a small amount of free and clear space. For example, the use of the SilverFit rehabilitation system in centres has been reported [26]. However, the system required the presence of a physiotherapist and large dedicated free space 5x5m; and as a result would not be suitable for use in most homes. Also, given a patients injuries from their fall or knee surgery, it would not be sensible for them to start moving equipment or furniture to create free space. Another issue with regard to knee and falls patients is the need to use supports which may be between the user and the cameras while performing the exercises. For example, a knee patient must sit or lie on a bed or chair in a comfortable position while performing the exercises [27-28], therefore occlusion would occur and the system would not show the users movement correctly.

More recently, researchers have started to explore the fusion of cameras systems and inertial motion sensors e.g. monitoring of the upper arm during eating activity after stroke [29] in order to overcome the limitations of camera based systems in handling occlusion [30]. Doyle et al. demonstrated the acceptance and usability of a home-based visualization system that uses 2 step counters and a web camera to track 3 reflective markers for a home physiotherapy program for falls prevention for seniors [18]. However, these setups would be unsuitable for providing precise movement information, which is important in maximizing physiotherapy benefits of the knee rehabilitation exercise program, although they were able to monitor the correct completion of exercises and track the number of repetitions. As with other camera based systems, the markers can easily be occluded and the cameras require sufficient free space around the user thus limiting the type of physiotherapy exercises that can be monitored. In this paper we present our visualization system that is capable of capturing finer motor movements without space constraints for two rehabilitation scenarios i.e. falls prevention and knee replacement rehabilitation.

3 Technology

We organized a series of user design workshops actively involving older adults in the initial design of the visualization tools [15], [31]. We also consulted falls experts and physiotherapists in order to obtain their opinions on the factors to consider in the delivery of useful feedback and advice to fallers and knee replacement patients undergoing home rehabilitation. It was important for us to identify the way in which the visualizations could improve the rehabilitation process, as well as the challenges that elderly users may have when using them and the novel technology. In this section, we describe the visualization tools and technologies that were evaluated.

3.1 Inertial Sensors

We developed inertial sensors that could be easily worn by the users. An orientation sensor fusion algorithm was implemented using the embedded C language, to translate the readings of the individual inertial sensor components (accelerometer, a rate

Fig. 1. The inertial motion sensor

gyroscope and a magnetometer) and output the orientation of the sensor in world space. While the accelerometer and gyroscope were needed to calculate the 3-dimensional orientation of the sensor, the magnetometer was included to reduce the effects of 'sensor drift' (i.e. the gradual wandering of gyroscope values over time) on the accuracy of the readings. The sensor data is sent wirelessly to a data receiver attached to a computer that uses the data to feedback user movements through the visualization software. Other significant features of the sensors included: a large switch to make it easy for elderly users to switch the sensor on, and a bright LED light to let them know when it was switched on.

3.2 Visualization Tools (VT)

For this research, two variations of the visualization tool were developed for each of the musculoskeletal conditions being investigated because the main requirements for both were different. At a base level, however, the visualization tools for both falls and knee replacement were somewhat similar e.g. they were both designed to provide visual feedback to users during home exercise based on the individual requirements of the various conditions. These requirements are highlighted in Table 1.

Table 1. Visualization Tool (VT) Requirements

Requirement	Solution provided by the VT
1. Show the user how and where on the body the sensors should be placed	Diagrams and text instructions showing the users the correct procedure
2. Show the user how the exercise should be done (emphasize quality of motion)	a. An animated 'guide mannequin' highlighting the correct pace and range of motion for each exercise b. Text-based instructions describing important advice about each exercise
3. Show the user how many repetitions have been achieved during exercise	A text-based repetition counter showing the number of repetitions done, and how many more to do
4. Feedback to the user on their own performance during exercise	a. A user mannequin showing the users' movements as captured by the inertial sensors b. A colored graphic showing the quality of each range of movement while the user moves the affected leg (knee VT only) c. A colored icon for each repetition showing how well each repetition was done (knee VT only)
5. Show a chart of weekly progress	Charts showing the average number of completed exercises and the quality of the range of motion associated with the exercises (knee VT only)

The software was written in C++, QT (to allow for flexible scripting) and Ogre (a 3D animation engine) and at the base level displays movement information using a stick figure, as well as other text based information [31-32].

3.3 Falls Rehabilitation Visualization Tool

Our initial discussions with falls experts revealed that fallers vary widely in their ability to do the exercises prescribed. Therefore, they advised that the users should be able to do the exercises based on their individual abilities, as they would still be able to improve their muscle strength and balance. For this reason, the falls VT did not include visual targets for range of motion. Instead, the users were encouraged to compare their movements as shown by the user mannequin to the guide mannequin in order to ensure that they did the movements correctly (Fig 2).

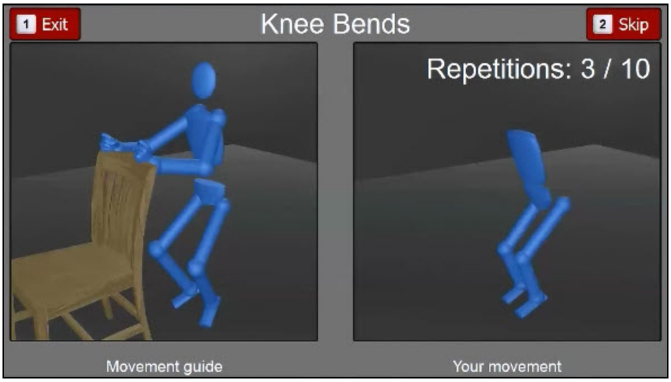


Fig. 2. Falls Visualization Tool: Knee Bend Exercise. Guide mannequin on the left and real time visual feedback on the right

We required a minimum of 6 sensors to animate the human model accurately for falls, as each major body segment tracked required a sensor. Whenever possible we try to minimize the number of sensors a user must wear, however six was the minimum number of sensors needed in order to capture and then visualize the user's movement in a usable way for the end user for falls. These sensors were strapped to the legs, chest, and pelvis in order to provide the feedback necessary to show the correct movements for all of the exercises. Navigation around the software was done through the use of keyboard presses only (key '1' was used to exit to the main menu; key '2' to go back to a previous exercise or skip instructions; and key '3' to go to the next exercise). Keyboard sticker labels were placed on these keys in order to make it easier to find these keys.

3.4 Knee Replacement Visualization Tool

The aims of knee replacement rehabilitation are to recover as full a range of motion as possible for the replaced joint and recover muscle strength. Following discussions with orthopaedic surgeons and physiotherapists, we therefore included targets for the users to achieve in the knee replacement VT. However, an overlaid ghost movement was not used because it could give rise to a situation whereby a user in the early stages of recovery could over extend or flex while trying to keep up with the range of movement and pace which can lead to injury. Another unique feature of the knee rehabilitation VT is that it gives real-time feedback to only the operated leg although the whole body is shown – to provide guidance on correct posture. This focus on the operated leg allows us to reduce the number of sensors needed to two – one above and one below the knee. This allows for an easier setup in the home. A graphic with a gradient of colors (red to green) was included to show the users the range of motion achieved during each repetition: a deep green represented the recommended range of movement; a light green represented an 'almost there' range; a yellow indicated a minimum range and a red indicated a poor range. For instance, Fig. 3 shows a typical visual feedback.

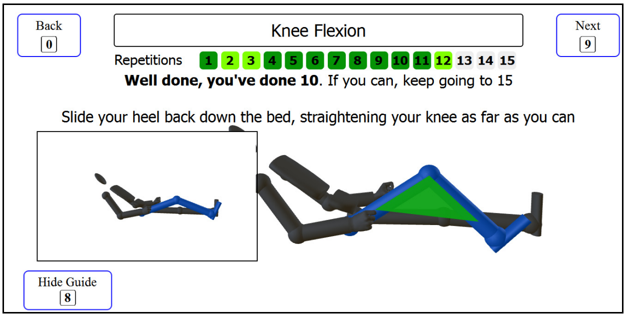


Fig. 3. Knee Replacement Visualization Tool: Knee Flexion Exercise. Guide mannequin in inset and real time visual feedback in the background

The colored repetition icons were also included to show how well each repetition was done. To keep a user motivated, messages such as "Well done, you've done 10, try to keep going to 15" is shown under the repetition counter. Unlike the falls VT, we opted for a simpler way to show the guide mannequin by making it show only when the user needs it. This way we reduced the cluster on the screen. The users interacted with the knee replacement VT using a remote control similar to a TV remote to facilitate ease of use. The visualization buttons and options were mapped to single numeric keys on the remote control to ensure that the VT can easily be manipulated from a comfortable position.

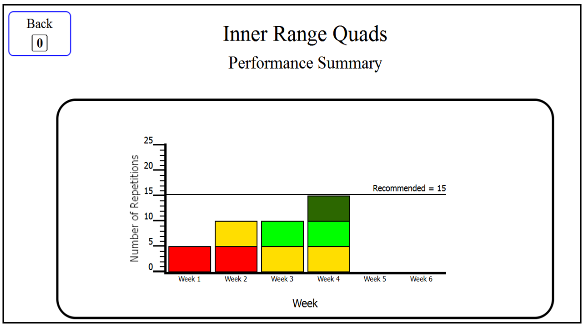


Fig. 4. Knee Visualization Tool: Weekly Progress Charts for Inner Range Quad Exercise

Progress charts were included in the knee replacement VT in order to track the users' quality of motion in the affected limb over an extended period of time (Fig 4). The color-coded bar chart shown in Fig.4. shows the quality of the repetitions recorded. This way, the user and the Allied Health Profession (AHP) could tell how much progress they had made with each exercise in terms of the ability to move the affected joint and the average number of repetitions completed per week. Both the target user and the AHP participants in our design workshops suggested that progress should be shown on a weekly basis rather than a daily basis because not every exercise session will show progress. AHPs will have remote access to VT midway into the rehabilitation period (clinical follow up) and if there is consistent deterioration in performance or number of repetitions, then urgent medical attention will be provided.

4 Methodology

Four usability studies were carried out with fallers and former knee replacement patients in the laboratory and in the home – in order to answer the following research questions:

- How easy is it for the users to use the motion sensors (strapping on and operating)?
- Does the use of visualizations enable the user to perform the rehabilitation exercises more correctly in comparison to the standard care exercise booklets?
- Are these visualization tools acceptable and usable by seniors?

The laboratory studies were undertaken first because we could control the environment e.g. setup, layout, interruptions and other health and safety variables. Once we had undertaken the laboratory studies and had eliminated any design issues that caused the users concerns we proceeded to the home studies.

4.1 Participants

University ethics was obtained for the studies. We recruited seniors aged 60 years and above for all four of the usability studies. Some of the participants were recruited by telephone, or had attended the previous design workshops we had undertaken one year before. Written consent was obtained from each of our participants. For the knee replacement studies, all of the users had their knee replaced in the past 18 months. For the falls studies, the participants had all had at least one fall in the12 months prior to the usability studies.

Table 2. Distribution of Participants

Group	Laboratory	Home
Fallers	3 participant (1 female, 2 males, mean age 68)	2 participant (1 female, 1 male, mean age 79)
Knee Replacements	3 participants (3 males, mean age 71)	3 participants (1 female, 2 males, mean age 63)

Their educational backgrounds were varied and so was their experience with computers – e.g. out of the 11 participants, 3 had never used a computer. Table 2 shows the distribution of participants.

4.2 Procedure

In the laboratory studies, the furniture was arranged in the laboratory to simulate a home environment with all the items necessary to carry out home exercises – a table to place a television and a laptop computer on; and a chair to do the exercises with (Fig 5). The participants were encouraged to think aloud while doing the tasks given.

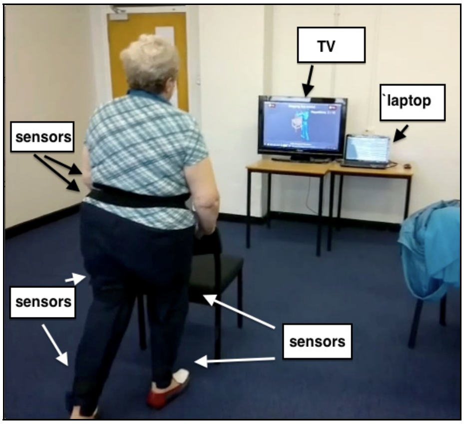


Fig. 5. Laboratory Setup: A participant using the Fallers VT for an exercise requiring the use of a support

The second sets of studies were then carried out in users' homes in order to evaluate these visualization tools in their intended environment (Fig 6).

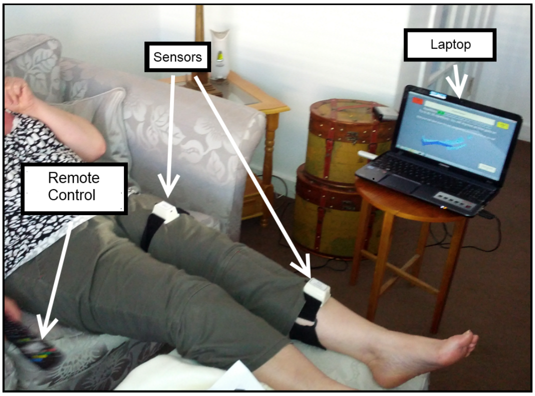


Fig. 6. Home Setup: A participant using the Knee VT

In both the falls and knee replacement studies, we used a within-subjects design because we believed that using the same participant to test the usability of both the exercise booklet and visualization would produce a more accurate comparison as each of the participants would have experienced the two types of rehabilitation aid first hand. Another merit of the within subject test design is that we can maximize the potential of our outcomes given the small number of participants.

The main procedure for the studies was divided into phases, each containing tasks, which were intended to investigate various aspects of the use of both standard care as well as the visualization tools. These phases are shown in tables 3 and 4 below:

Table 3. Main phases in the Falls studies

Phases	Procedure
Phase 1	Participants were asked to perform 3 exercises using the booklets (the exercises were the same for all of the participants)
Phase 2	Participants were asked to put on the sensors using instructions from the visualization tool only
Phase 3	Participants were asked to perform 3 exercises using the visualization tool (the same exercises done in phase 1)

Table 4. Main phases in the Knee Replacement studies

Phases	Procedure
Phase 1	Participants were asked to perform 3 exercises using either the booklet or visualization (the exercises were different for all of the participants)
Phase 2a	A short training session was provided to the participants on how to use the visualization tool
Phase 2b	Participants were asked to put on the sensors using instructions from the training, user manual or the visualization tool
Phase 3	The participants were asked to perform 3 exercises using the visualization or the booklet (same exercises done in Phase 1 but with a different tool).

Given the factors above, a qualitative and quantitative approach was used to develop the procedure involving the different tasks and short breaks were given at the end of each phase – each phase lasted between 10 and 20 minutes. At the end of each phase, participants were asked to complete short questionnaires, and semi-structured interviews on the tasks undertaken.

In phase 1, the participants were asked to do a few exercises with the booklet, while being observed. Two exercise booklets commonly used for home rehabilitation were used: a) The ‘Age UK’ exercise booklet for the falls studies [28] and b) The Joint Pathways knee rehabilitation exercise booklet for the knee replacement studies [27]. In phase 2, the participants were asked to follow instructions on the screen while they strapped the sensors onto the various limb segments. We wanted to investigate how easy the task was for them, as well as how well the instructions described the management of these sensors. We also noted how long it took them to put on the sensors. In phase 3, the participants were asked to do some exercises with the visualization tool, and provide feedback on their experiences with these tools.

The sensor calibration procedure used in the falls studies required the users to stand upright for 5 seconds before each exercise. An extra 'training' phase was included in the knee replacement studies because there was a more complex sensor calibration procedure used – than the one used in the falls studies –this was because for knee replacement we need a reference angle to provide accurate feedback on the knee's range of movement. The calibration procedure included putting a foam wedge with a known angle under the knee.

In the falls studies, the participants did the exercises with the booklet first before using the visualization tool. There were two main reasons for this – first, the participants had all used a booklet in the past for falls rehabilitation. Secondly, it was expected that because the visualization tool contained animated movements that demonstrated recommended pace – unlike the booklet – using it before the booklet could affect the way the participants carried out an exercise using the booklet. As regards the participants in the knee replacement studies not all of them had used the knee replacement booklet in the past. Therefore, to minimize the potential transfer of learning by using either the exercise booklet or visualization first, the order of the use of the two tools was randomized among the participants with 3 of the participants using the booklet first while the other 3 used the visualization tool first.

5 Key Observations and Findings

In both the falls and knee studies, a facilitator closely observed the participant as the participant carried out the tasks in the various phases. In addition we took notes and timed the users during the tasks. In this section, we report on our quantitative and qualitative findings for both the home and laboratory tests because the results were broadly similar.

5.1 Performance of Rehabilitation Tools

Falls Rehabilitation

We observed that with the standard exercise booklet the participants performed the repetitions in each exercise at a quick, but roughly uniform pace. With the visualization tool (VT), the repetitions were initially quicker (during the first few repetitions), and then slower and more regular afterwards. Figure 7 shows the difference in average time per repetition for the 'knee bends' exercise. The results were similar for the other exercises, with the participants performing the repetitions more slowly and in a more controlled manner while using the visualization tool.

When asked to comment on using the visualization tool for their exercises, the participants' opinions were similar across both the laboratory and home studies. The main responses were that:

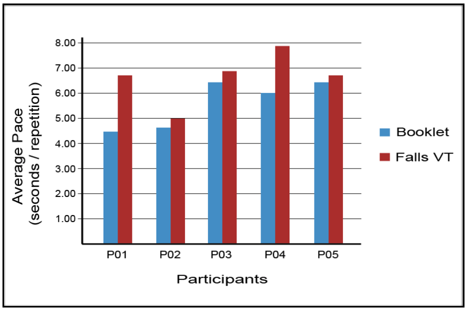


Fig. 7. Average Repetition-pace for the Knee Bends exercise

1. They tried to follow the movements as shown by the guide mannequin; the result of this was that they started to slow their movements down after a few repetitions.
2. They concentrated on the guide mannequin and not on the user mannequin i.e. the mannequin that showed their own movement, for most of the time. However when we further explored this observation with them, they said they would start to concentrate more on their own movement feedback as their confidence and understanding of the system grows.
3. They liked the ability to see the user mannequin that showed their own movement.
4. They all felt that they would use both the booklet and visualization tool if given the chance, but would use the visualization tool more, as they found the feedback helpful and important to improving their confidence in the exercise program.

Knee Replacement (KR) Rehabilitation

The findings were similar for the knee studies in terms of quality of movement, with the visualization tool encouraging slower, more controlled movements. Fig. 8 shows the Straight Leg Raise exercise and the results shown here shows a typical result for all of the other exercises tested.

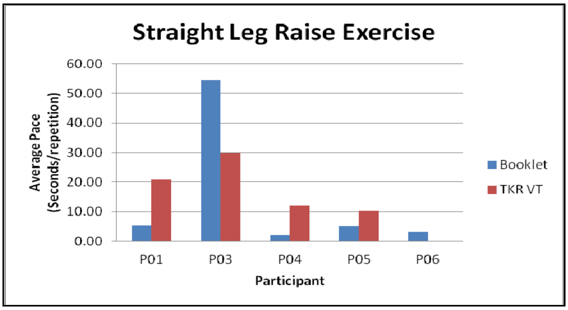


Fig. 8. Average pace for the Straight leg raise exercise

One issue that we observed was with the booklet instruction for the Straight Leg Raise, the instruction asks the patient to raise your leg 6 inches. All of the participants had different concepts of how high 6 inches was. We noted that participants P01, P05 and P06 had not used the exercise booklet before and participant P06 couldn't complete the exercise with the visualization tool because there was a software bug in the calibration algorithm. As the participants used the booklet, the exercises were performed faster than the recommended pace except for P03 who did the exercise very slowly because he was very talkative and was distracted by waiting for responses from the facilitator. Furthermore, none of the participants managed to keep the hold time for the exercises and lost count of the number of repetitions completed.

On the other hand, we observed that while using the visualization tool, the participants concentrated on their own movement and tried to achieve the green targets. One of the participants in the knee replacement studies forgot his reading glasses at home however, he was still able to use the tool as large font sizes had been used. Another participant brought the laptop computer closer so it could be seen. However, she commented that the elements on the screen could be seen clearly.

When asked to comment on using the visualization tool for their exercises, the participants' opinions were similar across both the laboratory and home studies. The main responses were that:

1. They liked the color-coded fan concept and thought that it helped them to achieve the correct performance of the exercise. One participant commented: *"It is very useful to have fullness of each exercise calibrated by color code"*.
2. Of particular note is P01 who got the instruction for the straight leg raise exercise wrong and was doing something different. He couldn't get into the green zone. He commented as follows *"I'm getting this wrong isn't it? I will show the guide. Oh! I was wrong"*. He was able to correct this without any cue from the facilitator when he viewed the guide mannequin.
3. When asked, *"To what degree did the visual feedback help you to perform the rehabilitation exercises?"* 5 out of the 6 participants found it very helpful and one participant found it moderately helpful. A typical comment received was *"This is a big help because you don't count seconds when doing the static quad on your own. Sometimes you wait too long or too short"*.
4. All of the participants successfully accessed the progress charts. See Fig. 4. When the participants were asked, *"To what degree do you think the progress charts will help you during your home rehabilitation?"* Using a Likert scale of 1 – 5 with 1 being 'Not helpful' and 5 being 'very helpful'. 2 out of 3 participants said 'very helpful' and 1 participant said 'helpful'. One participant commented, *"It is very important to stimulate motivation to continue at an appropriate level"*. All of the participants found the charts very easy to interpret.

5.2 Evaluation of the Sensors

We took into account that the number of sensors required for the falls and knee replacement visualization tools are 6 and 2 respectively. We asked the participants to put on the sensors using the instructions provided on the screen. They were also asked to repeat this task after five minutes (in order to see how well they could remember to put on the sensors on the second attempt). Most of the participants were able to put the sensors on correctly on the first and second attempt. Some of the participants in both falls and knee replacement studies put the first sensor on the wrong way (upside down). However, they corrected this when they proceeded to the exercise session and saw a reminder to make sure that the switch was facing upwards, and that they could see the red LED light.

When asked to rate how easy or difficult this task was, 5 of the participants said it was very easy and 4 were confident that they could set up the visualization tools on their own in the home. It should be noted (see Figure 2) that the sensors had been labeled and had to be worn on the front of specific limb segments with the switch facing up. When asked about "how comfortable it is to wear the sensors", all the participants said the sensors were very comfortable to wear for long periods. All the participants also said they were able to follow the on screen instructions to successfully complete their exercises.

The sensors performed well in both the home and laboratory environment but for one participant in the Knee VT group. The reason was that the calibration algorithm relied on facing the north direction and the participant's living room was facing the East, we have since eliminated the requirement for a participant to face in a particular direction.

5.3 Interaction with the Visualization Tools (Keyboard and Remote Control)

We investigated 2 setups; the falls visualization used the computer keyboard to navigate through the different visualization options while the knee replacement visualization tool provided two options the keyboard or a remote control. We introduced the remote control as a controller of the tool because of the limited movement of the users in the early stages after knee replacement surgery.

All the participants found navigating the visualization tools using the keyboard easy as the navigation keys were clearly labeled and visible. In the knee replacement studies, the participants preferred the use of the remote control to the keyboard because they didn't have to stand up from the exercise position to operate the tool. All of the participants were able to navigate through the menu items with ease.

6 Discussion

6.1 Home Exercise (Booklet vs. Visualization Tool)

One of the objectives of the user studies described in this paper was to investigate whether two novel visualization tools could improve the quality of home rehabilitation for falls and knee replacement – compared with the use of standard care. Based on our findings, we identified some issues with the use of booklets (standard care), which makes them less than an ideal solution for effective therapy during rehabilitation.

First, the participants often completed the exercise repetitions too quickly with the booklet. This means that they do not get the therapeutic benefit of a controlled movement. By using the visualization tools, they were able to avoid this problem, as the participants followed the movements as demonstrated by the animated mannequins; and were able to achieve the recommended quality of movement.

Secondly, the participants missed some repetitions with the booklet because they had to rely on their memory to keep track of the repetitions. Both the quality and the quantity of physiotherapy exercises received influence the recovery of function, range of movement and muscle strength. This did not happen when they used the booklet but did happen when they used the visualizations, as they were guided through each repetition and there was a repetition counter.

Thirdly, the visualizations showed the participants their movements during each exercise (something that the booklet cannot do) and this gave them more confidence in the exercise programme.

Fourthly, showing two contrasting animations side by side resulted in the users focusing on one and missing out some frames in the other. This in the rehabilitation context is not a bad thing as long as the most important window is emphasized and held in focus. For the Falls VT, it was the correct movement pattern that was important which the users rightly tried to imitate while for the Knee VT the more important image was bigger than the guide mannequin and had changing color graphics that users held in focus. The ability to turn off the guide mannequin was also a useful future that helped users concentrate on a particular window.

Fifthly, the ability to see progress over time was considered a major motivating factor in continuing with the home exercise programme. In the past, the people had used subjective means to measure their progress – e.g. amount of pain in the joint, the ability to stand up quicker, completion of exercise diaries, etc. However, these are subjective and cannot easily be tracked over time. Having objective information about progress over time could provide insights into the effectiveness of standard physiotherapy exercises. In addition, anxiety regarding recovery could be reduced if a user could objectively see that they are capable of greater movement.

6.2 Usability of Visualization Tool

Our findings suggest that seniors are capable and able to use technology for their home rehabilitation. In this section we highlight some of the design issues found during our studies and how these could contribute to our understanding of the issues surrounding acceptance of technology for rehabilitation in the home.

Convenience and ease of setup: most seniors are used to operating technologies like the television. Using this metaphor in the development of home rehabilitation technology enabled seniors to feel more confident in their abilities to use a new technology. An acceptable technology should take only a couple of minutes to start and the menu easily accessible via limited numeric and arrow keys. The majority of the participants were able to set up the exercise sessions and correctly calibrate the sensors using the instructions in the visualization tools.

Initial Training and clear onscreen Instructions: An initial training session in the home environment that includes a practical walkthrough of how to use a new visualization technology increases the likelihood of its acceptance. Furthermore, because some seniors are not familiar with some technical jargon such as USB these terms should be avoided. If these terms are to be used in such a tool, it should be explained in a straightforward way so that the users understand it better. Memory is known to decline with increasing age. Therefore, it is important to provide instructions that are able to give the users the right information as soon as possible.

Accessibility: It was important for us to ensure that the visualizations were clear and visible, taking into account eyesight problems – which are common in the elderly population. As highlighted in the findings, the use of larger font sizes, blank spacing, red, yellow and green colors in close proximity was well received - seniors appreciate bright and vibrant colors most especially when it provides information about correct performance of an objective.

7 Conclusion and Future Work

The main function of musculoskeletal rehabilitation exercises is to recover impaired motor function in muscles and joints in the lower body. In the home setting, the lack of adherence to rehabilitation could have serious implications for the health and independence of seniors. Furthermore, if the recommended quality of motion (range and pace of movement) is not observed for these exercises, the seniors will probably not obtain the full recovery benefits offered by the rehabilitation programme(s). In this work, we evaluated two novel visualization tools with fallers and knee replacement patients; and compared the use of these tools to standard rehabilitation care (booklets) in the laboratory and in the home. One of the main objectives of this research was to investigate how useful the visualization tools were (versus standard care) in encouraging the recommended quality of motion for the exercises – thereby promoting the maximum therapeutic value of these exercises.

Based on our findings we argue that by using booklets to exercise, seniors may not observe the recommended quality of movement. The main findings supporting this are a) the inability to observe the correct pace of movement from static diagrams and b) the seniors could not tell whether they were doing the exercises the correct way. By using the visualization tools, the seniors were able to overcome these limitations, thereby getting the most out of the exercises from a therapeutic standpoint. The seniors felt that the ability to observe the correct movements – as well as receive feedback on their movements and progress – during rehabilitation was essential to increasing their confidence and motivation to exercise. Furthermore, based on the feedback from our participants, these visualization tools were easy to use and acceptable for home rehabilitation use. Another objective of this study was to investigate the usability of the body-worn sensors necessary to interact with the visualization tools. Our findings suggest that the seniors found the sensors easy to use and that they were confident in their ability to use and manage (e.g. placement on the body, charging, etc.) them without the assistance of a professional.

The study results are limited by the number of participants and the amount of time that the participants had to interact with the visualization tools. The effectiveness of these visualization tools in improving adherence to home rehabilitation programmes over an extended period of time still needs to be assessed. This is the basis for our future work, as we are now deploying these tools in home based pilot randomized controlled studies for 6 -12 weeks.

Acknowledgements. Funding for this study comes from the Medical Research Council, UK, LLHW (Phase 2), Grant ID: 91021.

References

1. Woolf, A.D., Pfleger, B.: Burden of major musculoskeletal conditions. Bull. World Health Organ. 81, 646–656 (2003)
2. Scottish Arthroplasty Project Biennial Report 2012, NHS National Services Scotland, p. 9, http://www.arthro.scot.nhs.uk/Reports/sap_national_report_2012.pdf
3. Skelton, D., Todd, C.: What are the main risk factors for falls among older people and what are the most effective interventions to prevent these falls? Copenhagen, WHO Regional Office for Europe (2004), http://www.euro.who.int/document/E82552.pdf
4. Thomas, S., Mackintosh, S., Halbert, J.: Does the 'Otago exercise programme' reduce mortality and falls in older adults?: a systematic review and meta-analysis. Age Ageing 39(6), 681–687 (2010)
5. Applegate, W.B., Miller, S.T., Graney, M.J., Elam, J.T., Burns, R., Akins, D.E.: A Randomized, Controlled Trial of a Geriatric Assessment Unit in a Community Rehabilitation Hospital. New England Journal of Medicine 322, 1572–1578 (1990)
6. Medical Advisory Secretariat: Physiotherapy rehabilitation after total knee or hip replacement: an evidence-based analysis. Ontario Health Technology Assessment Series 5(8) (2005)
7. Nyman, S.R., Victor, C.R.: Older people's participation in and engagement with falls prevention interventions in community settings: An augment to the Cochrane systematic review. Age Ageing (2011), http://www.ncbi.nlm.nih.gov/pubmed/21875865
8. Pisters, M.F., Veenhof, C., Schellevis, F.G., Twisk, J.W.R., Dekker, J., De Bakker, D.H.: Exercise adherence improving long-term patient outcome in patients with osteoarthritis of the hip and/or knee. Arthritis Care & Research 62, 1087–1094 (2010)
9. Scott, J.C.: Osteoporosis and hip fractures. Rheumatic Diseases Clinics of North America 16(3), 717–740 (1990)
10. Fessel, K., Nevitt, M.: Correlates of fear of falling and activity limitation among persons with rheumatoid arthritis. Arthritis Care and Research 10(4), 222–228 (1997)
11. Button, K., Iqbal, A.S., Letchford, R.H., Van Deursen, R.W.M.: Clinical effectiveness of knee rehabilitation techniques and implications for a self-care treatment model. Physiotherapy 98, 287–299 (2012)
12. Wheeler, J.W., Shull, P.B., Besier, T.F.: Real-Time Knee Adduction Moment Feedback for Gait Retraining Through Visual and Tactile Displays. Journal of Biomechanical Engineering 133, 041007 (2011)
13. Walker, C., Brouwer, B.J., Culham, E.G.: Use of visual feedback in retraining balance following acute stroke. Phys. Ther. 80, 886–895 (2000)
14. Zhou, Hu, H.: Human motion tracking for rehabilitation—A survey. Biomedical Signal Processing and Control 3, 1–18 (2008)
15. Uzor, S., Baillie, L., Skelton, D., Fairlie, F.: Identifying barriers to effective user interaction with rehabilitation tools in the home. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) INTERACT 2011, Part II. LNCS, vol. 6947, pp. 36–43. Springer, Heidelberg (2011)

16. Suttanon, P., Hill, K.D., Said, C.M., Byrne, K.N., Dodd, K.J.: Factors influencing commencement and adherence to a home-based balance exercise program for reducing risk of falls: perceptions of people with Alzheimer's disease and their caregivers. International Psychogeriatrics 24, 1172–1182 (2012)
17. Lim, C.K., Chen, I.-M., Luo, Z., Yeo, S.H.: A low cost wearable wireless sensing system for upper limb home rehabilitation. In: 2010 IEEE Conference on Robotics Automation and Mechatronics (RAM), pp. 1–8 (2010)
18. Doyle, J., Bailey, C., Dromey, B., Scanaill, C.N.: BASE - An interactive technology solution to deliver balance and strength exercises to older adults. In: 4th International Conference on Pervasive Computing Technologies for Healthcare, PervasiveHealth, pp. 1–5 (2010)
19. Harms, M.: Advancing technology in rehabilitation. Physiotherapy 98, 181–182 (2012)
20. Tanaka, K., Parker, J., Baradoy, G., Sheehan, D., Holash, J.R., Katz, L.: A Comparison of Exergaming Interfaces for Use in Rehabilitation Programs and Research. LDG 6, 69–81 (2012)
21. Gerling, K.M., Schulte, F.P., Masuch, M.: Designing and evaluating digital games for frail elderly persons. In: Proc. of ACE 2011 (2011)
22. Borghese, N.A., Pirovano, M., Mainetti, R., Lanzi, P.L.: An integrated low-cost system for at-home rehabilitation. In: 2012 18th International Conference on Virtual Systems and Multimedia, VSMM, pp. 553–556 (2012)
23. Theng, Y.-L., Dahlan, A.B., Akmal, M.L., Myint, T.Z.: An exploratory study on senior citizens' perceptions of the Nintendo Wii: the case of Singapore, pp. 1–5. ACM, Singapore (2009)
24. Fung, V., Ho, A., Shaffer, J., Chung, E., Gomez, M.: Use of Nintendo Wii Fit™ in the rehabilitation of outpatients following total knee replacement: a preliminary randomised controlled trial. Physiotherapy 98, 183–188 (2012)
25. Ziefle, M., Röcker, C.: User Diversity as a Challenge for the Integration of Medical Technology into Future Smart Home Environments. In: Human-Centered Design of E-Health Technologies, pp. 95–126. IGI Global, USA (2010)
26. SilverFit, `http://www.silverfit.nl/en.html`
27. JointPathways™ Patient Guides for Knees, Golden Jubilee National Hospital NHS Scotland
28. Age UK. Preventing Falls: Strength and balance exercises for healthy ageing (2011), `http://www.ageuk.org.uk/Documents/EN-GB/ID8950%20Strength%20And%20Balance%20Book.pdf?dtrk=true`
29. Hondori, M.H., Khademi, M., Lopes, C.V.: Monitoring Intake Gestures using Sensor Fusion (Microsoft Kinect and Inertial Sensors) for Smart Home Tele-Rehab Setting (2012), `http://www.isr.uci.edu/~mkhademi/files/Publications/HIC2012.pdf`
30. Bo, A.P.L., Hayashibe, M., Poignet, P.: Joint angle estimation in rehabilitation with inertial sensors and its integration with Kinect. In: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, pp. 3479–3483 (2011)
31. Uzor, S., Baillie, L., Skelton, D.: Senior designers: empowering seniors to design enjoyable falls rehabilitation tools. In: Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems, pp. 1179–1188 (2012)
32. Loudon, D., Macdonald, A.S., Carse, B., Thikey, H., Jones, L., Rowe, P.J., Uzor, S., Ayoade, M., Baillie, L.: Developing visualisation software for rehabilitation: Investigating the requirements of patients, therapists and the rehabilitation process. Health Informatics Journal 18, 171–180 (2012)

Walking in the Wild – Using an Always-On Smartphone Application to Increase Physical Activity

Tim Harries[1], Parisa Eslambolchilar[2], Chris Stride[3], Ruth Rettie[1], and Simon Walton[4]

[1] Business School, Kingston University, Kingston Upon Thames, UK
[2] Computer Science Department, Swansea University, UK
[3] Institute of Work Psychology, Management School, The University of Sheffield, UK
[4] Oxford e-Research Centre, University of Oxford, UK
{T.Harries,R.Rettie}@kingston.ac.uk, csparisa@swansea.ac.uk, c.b.stride@sheffield.ac.uk, simon.walton@oerc.ox.ac.uk

Abstract. This multidisciplinary paper reports on a large-scale field trial, designed and implemented by a group of social scientists, computer scientists and statisticians, of a new smartphone-based app for the promotion of walking in everyday life. The app, bActive*,* is designed for a more diverse range of users than the typical active-lifestyle app, since it requires neither additional equipment nor a great deal of commitment to exercise. As a result, it can raise awareness of walking and promote walking amongst those with only a casual or hesitant engagement with the topic. The 6-week randomised controlled trial with 22-40 year-old male participants (N=152) indicates that bActive prompted users to increase the amount of walking they did by encouraging them to value and increase walking that is incidental to normal everyday activities. Longitudinal data analysis showed that use of the app increased walking by an average of 64% but did not find any evidence to suggest that the inclusion of comparative social feedback improves the impact of such apps on male participants.

Keywords: walking, feedback, norms, app, active-lifestyle, social sharing.

1 Introduction

Walking is generally considered enjoyable, relaxing [32], beneficial for general health [2] and helpful for the prevention of obesity and chronic disease [28]. It is "readily repeatable, self-reinforcing and habit-forming" [31] and is the most widely accessible type of exercise because it is inherently safe, requires no special skills, location or equipment and can easily be included in domestic and work routines [32]. However, most residents of advanced economies take far less than the daily total of 10,000 steps generally recommended for good health [5,10,20,29,33].

Unlike most previous apps aimed at increasing physical activity, the one described here, bActive, requires no special equipment and does not need to be activated by the

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 19–36, 2013.

user in order to track activity. This makes it easier to acquire (users only have to download the app and do not need to purchase any physical device), easier to use and more suitable for those with a passing or tentative commitment to becoming more active. The only demand made by bActive is that users engage with the feedback. This is encouraged firstly by their being able to do so at their convenience and, secondly, by the inclusion of trend data that is likely to engage their interest.

The bActive app collects data from the phone's built-in accelerometer and utilises an always-available display to deliver information to users about the number of steps they have taken. The presentation of this information was designed to increase walking activity in three principal ways. First, users are made more aware of the exercise involved in purposive 'walks' and that inherent in the walking involved in day-to-day activities such as shopping and work. Second, users are able to track their own activity over time. Third, they can be offered the chance to compare their activity levels with the average activity levels of others.

This study compared a control condition (no feedback) with the use of feedback limited to a user's own walking and the use of comparative data i.e. a *social norms approach* [11,18]. The social norms approach has been successfully used in fields as diverse as alcohol abuse, sexual behaviour, the payment of tax debts and domestic electricity consumption [9] but has not previously been applied to exercise or delivered by a smartphone app. The approach has two main elements. First, on the assumption that forces of conformity encourage people to emulate social norms, it provides individuals with information about the average behaviours of a group of salient others (the *descriptive norm*). Second, to avoid encouraging change in a negative direction (e.g. regression towards a lower activity level), some social norms practitioners provide users with moral approval for 'good' behaviour (the *injunctive norm*) [12,34].

A meta-analysis of the effectiveness of this approach in its main area of application, alcohol abuse by students, is presented by Bosari and Carey [6]. In this domain, the approach usually begins with survey research on respondents' own alcohol consumption and their assumptions about that of other students in the same university. Where this reveals a tendency to overestimate alcohol consumption, the discrepancy is conveyed to the student population of the university via poster campaigns. In another example, a randomized controlled trial of a US programme that posted reports containing social norms information with households' electricity bills found that this intervention reduced consumption by around 2% [1].

2 Related Work

Pedometers use piezoelectric accelerometers to count the number of steps walked and can be worn on the body or carried in a pocket. Pedometers have long been of interest for their ability to encourage more active lifestyles [3,16]. However, exercise promotion programmes have typically used pedometers alongside other resource-intensive activities such as classroom training and face-to-face sessions [7,23], so little empirical data exists regarding the use of pedometers in more natural contexts.

Pedometers exist either as devices dedicated to the measurement of exercise, or as embedded features in other equipment such as mobile phones. This difference has implications for accuracy and usability. Accuracy is highest amongst dedicated pedometer-devices, where it exceeds 96% at speeds of over 3 miles per hour (mph), dropping to 74%-91% at 2-3mph and 60%-71% at below 2mph [3]. However, users have to be committed enough to fitness and lifestyle-change to purchase the pedometer and remember to wear or carry it. This, and questions of fashion, design and convenience, can deter some people from using such devices [13]. In contrast, mobile-phone pedometer apps have the advantage of being embedded in equipment that people already own and keep on their person and their in-built display and communication capabilities allow users to share their feedback with others. Hence, smartphones are increasingly being used to address the problem of sedentary lifestyles [15].

2.1 Goal Setting

Goal-setting and performance feedback, which most generations of smartphones make convenient and easy,[1] are important influences on individual behavior [26] and are generally considered key features of technologies intended to encourage physical activity [13].

A number of factors influence the effectiveness of a goal. Firstly, it must be accepted by the individual and not in conflict with their other goals [27]. Secondly, increased difficulty is said generally to increase motivation [27], possibly because of the greater potential self-satisfaction that accrues to the user on achievement of the goal [36]. However, goals must not be considered unachievable [27] and though failure to attain a goal can increase motivation [8], it can also have the opposite effect [19]. Thirdly, the discounting of future benefits leads to the argument that short-term goals are more effective than longer-term ones [36].

One approach to determining physical activity goals is demonstrated by Chick-Clique [38] and UbiFit Garden [14,15], in which users set their own daily step-count goals. However, this approach runs the risk that inexpert users will set goals that are either too difficult or too easy and that do not, therefore, provide optimal motivation [33]. In contrast, in Fish'n'Steps [25] goals are set at a modest level by automatically using baseline step-counts as reference points and taking the findings of previous studies as a guide to what users could reasonably be expected to attain. Furthermore, Fish'n'Steps breaks longer-term goals down into proximal, daily, sub-goals. No research has been done to directly compare these two, contrasting, approaches.

2.2 Social Sharing

Mechanisms that facilitate social influence are also often considered essential for devices that encourage physical activity [13]. A number of smartphone apps include

---

[1] See, for example, Nokia's Wellness Diary (`http://betalabs.nokia.com/apps/wellness-diary`)

this facility – e.g. [13,38,25]. In Fish'n'Steps, for example, each user is presented with a fish avatar whose growth, emotional state and behaviour reflect the number of steps the participant takes each day. The avatars of each group of users are displayed on a screen in an area shared by all its members (e.g. the social space in an office) and also on users' personal websites. A second example is MapMyWALK,[2] whose popularity is demonstrated by the fact that it had been downloaded from Android App Shop 250,000 times by July 2012. This app allows goals, routes, distances and walking speeds to be shared with friends and family members via email and social media.

Evidence on the effectiveness of these apps is mixed and small sample sizes cast doubt over its validity. For example, although Fish'n'Steps is reported to have caused some participants to increase their step-count, the unhealthy fish avatars that result from low activity levels caused others to drop out of the study. Similarly, although Consolvo et al [13] claim that social comparisons influenced their (all female) participants, the report of their three-week pilot of Houston with friendship groups of young females (N=13) reveals that the sharing of data did not have any significant impact on step counts. Finally, in a trial of Chick-Clique (N=7), group performance was reported by the participants (13-17 year-old girls) as the most "powerful method of changing behavior" [38; p1877], but the sample was too small to test this claim.

3 The bActive App

Although the design of the bActive app drew on the principles detailed in the previous section, it differs in three key ways from most of the apps described. Firstly, no assumptions are made about users' willingness to spend time and effort tracking their activity levels. Users simply have to download the app and carry their mobile with them in a trouser pocket. There is no requirement for additional equipment such as pedometers or foot pods, or for the data entry required for diarisation. While motivated users may be prepared to carry additional devices to measure specific physical activities, they are less attractive to those who are ambivalent about the benefits of measurement or their ability to become fitter and healthier [13], and hence are unlikely to promote ubiquitous use.

Secondly, unlike apps that activate only when users notify them at the start of an exercise event (e.g. MapMyWALK), bActive measures activity continually and without the need for any user action. Users are therefore rewarded with their activity data without having to make any initial effort or remember to switch the recording function on and off. This means, in addition, that rather than focusing uniquely on intentional exercise events such as hikes or walks, bActive also measures the exercise inherent in routine activities such as shopping, walking children to school or moving around at work. As a result, the emphasis is on the adoption of a healthy lifestyle, rather than on participation in walking as sport or recreation. This too is one of the

---

2 `http://www.mapmywalk.com/`

reasons for not using GPS in bActive, for much incidental walking (e.g. shopping- and work-related walking) occurs in geographically confined spaces and is therefore less amenable to measurement by GPS.

The third difference concerns goal-setting. In bActive, formal goal-setting, training and coaching elements are replaced by self-generated, informal targets that result from a user's engagement with the feedback information. As a result, rather than feeling that they are engaging in a formalized exercise program, users are allowed to respond to this information in whatever way they wish. As argued by Thaler and Sunstein [37], behavioral feedback forms part of the choice architecture that nudges behavior. In this case, the bActive feedback nudges users to walk more. The only action required of them is that they occasionally bring the app to the foreground by clicking on the bActive icon, and this is subtly prompted by the presence of the bActive icon on the phone screen (see Figure 1).

To engage people who are initially less committed to increasing their activity, it is particularly important that bActive is seen as interesting and fun to use. Learning from Fish'n'Steps and UbiFit Garden, it uses non-literal, light-hearted visual representations of behaviour. It also provides trending information (as in Fish'n'Steps, UbiFit Garden, Into and Houston); gives positive reinforcement (learning from UbiFit Garden and Houston and from the problems experienced by Fish'n'Steps), and, like Houston, Chick Clique, UbiFit Garden and Into, provides opportunities for users to reflect on their own activity. Finally, like the social gaming and social data sharing features of Fish'n'Steps, Houston, Chick-Clique and Into, the social norms information within bActive is designed not only to prompt increased walking, but also to encourage engagement with the feedback.

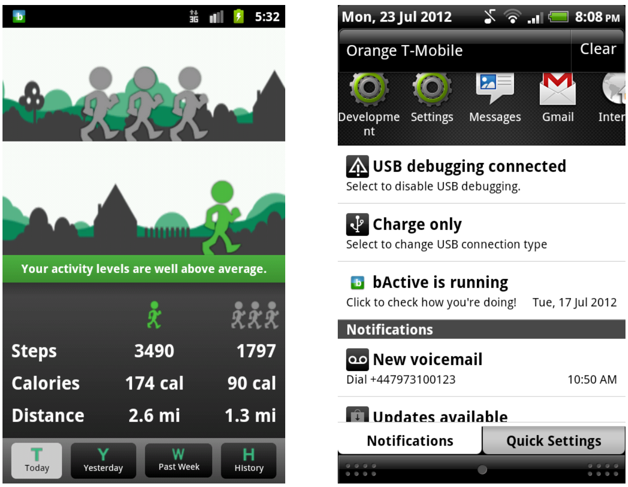


Fig. 1. Left: bActive's *Today* screen, as seen by those in the social feedback condition. (Note: the bActive icon is visible in the top left corner). Right: an entry on the Android notification bar indicates that the app is running in the background.

3.1 Design Philosophy

To test the efficacy of individual and social norms feedback, three different versions of the app were created: a ‘null’ version for participants in the control condition, which provided no user interface and gave no access to feedback; a ‘partial’ version for those in the individual condition, which displayed individual data only; and a ‘full’ versions for those in the social norms condition, which displayed both individual data and group averages. This section focuses on the version provided to those in the latter condition, for this was the most complex in terms of design.

Developed for Android 2.3, bActive incorporates automatic step counting alongside on-demand real-time and historic feedback of the number of steps taken by the user and the average of a group of other users. It also logs the frequency with which users open the app and the length of time the app is open on the display. The app is intended to be used on-the-go, so data clarity is emphasized and users are not asked to perform any retrieving or filtering tasks. To verify ease of use, the app was piloted twice and testing performed in a variety of outdoor conditions.

The app features four views of the data: *Today*, *Yesterday*, *Past Wee*k and *History* (Figure 1). The *Today* screen shows progress for the current day in terms of *steps*, *distance* and *calories* (calculated for the target demographic of young men) [17,39]. Values for the group average are displayed alongside those for the individual user. To facilitate rapid review of progress, an animated avatar representing the user (the green walking figure) is shown either behind, in front of or alongside an animated group of avatars (the grey walking figures) that represent the average activity for those in the comparison group. A banner just below the avatars displays a feedback message that varies according to how the individual’s performance compares to that of the group. For those above average, it toggles between a descriptive norms message (e.g. “Your activity levels are above average”) and an explicitly evaluative injunctive norms message (e.g. “☺☺ Well done, keep it up!”). If the individual’s activity is below average, the banner displays a single descriptive norms message (i.e. “Your activity levels are below average”). The *Yesterday* screen is identical to the *Today* screen, but gives the previous day’s results. The *Past Week* screen (Figure 2) displays a line graph depicting activity levels for the previous seven days, including averages for the group and for the most active 20%. Identical to the *Past Week* screen in format, the *History* screen allows users to swipe back and forth between different weeks.

3.1.1 Client/Server Architecture

The app uses the handset's mobile data connection to send users’ activity levels and app usage to a central MySQL, and retrieves the group’s activity average from that database. Step data is sent from the phone by a background Android service that transmits step data every two hours. This enables data to be updated asynchronously as users visit the screens, and ensures that they are shown up-to-date information. If lack of connectivity prevents data transmission, the app reduces the delay for activity transmission from two hours to one hour until a successful link is established. Data on use of the app is transmitted every four hours.

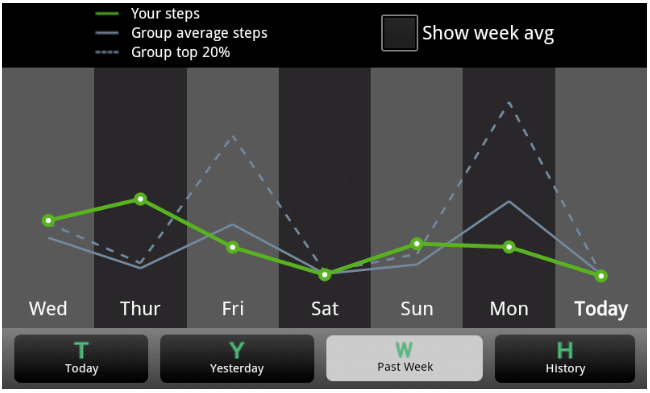


Fig. 2. The *Past Week* screen as seen by those using the full version of the app

A delay-based caching mechanism decides when particular elements of data on a user's device need to be updated and only requests fresh data when required. Meanwhile, the app provides the user with an indication of how much data it is using.

3.1.2 Activity Monitoring

The app's activity monitoring system is implemented as an Android service. Designed to be as autonomous as possible, it starts upon device boot and automatically recovers from unexpected crashes. On starting, it registers itself as a foreground service in order to prevent the Android process scheduler from hibernating the process when there are non-severe memory requests from other apps. An entry on the Android notification bar to indicate to the user that the app is running (Figure 1).

When the service is running, each accelerometer reading arriving as an (x, y, z) array is treated as a vector of magnitude $\boldsymbol{m}$. Instead of computing the true vector magnitude, the app simply computes m^2 as $x^2 + y^2 + z^2 - G^2$, where $\boldsymbol{G}$ is Android's accelerometer constant for Earth's gravity. The step counting algorithm is based on that of Mladenov and Mock [30], which treats accelerometer values as graph $\boldsymbol{y}$-values over time.

3.1.3 Keep-Alive / Battery Conservation Strategy

To preserve battery life, the Android power manager normally puts the CPU to sleep when the device screen is turned off through display timeout. Events such as phone calls and activated alarm clocks can turn the CPU and/or screen back on.

To ensure that the device captures a user's activity levels throughout the day without adversely affecting battery life, bActive uses a combination of two Android concepts: WakeLocks[3] and Alarms. Due to inefficient development of WakeLocks since Android 2.2,[4] a power strategy was used that allowed the CPU to sleep partially

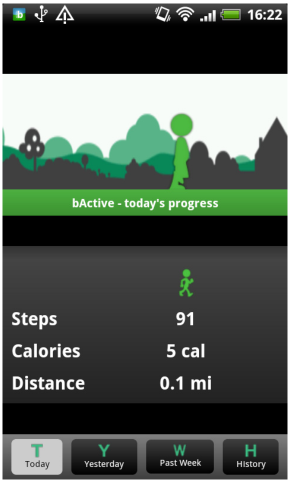
[3] A mechanism to prevent the device from entering a low-power state.

[4] `http://developer.android.com/reference/android/os/PowerManager.html`

during periods of inactivity rather than reading the accelerometer at all times. The Android Alarm Manager wakes the phone from a low-power state once every 30 seconds and takes around twenty measurements from the accelerometer at a low frequency. If any of the readings are interpreted as representing a step, the accelerometer frequency is set to high and the device continues monitoring user activity until one minute after the final step is detected. At this point, the WakeLock is released, allowing the device to sleep for a further 30 seconds. Although essential for power saving, this strategy has the disadvantage that short bursts of activity (e.g. walking around a kitchen while cooking) will sometimes be missed if they are dispersed across periods of inactivity. This strategy may not be required in future generations of Android OS if the current issue with the WakeLocks is addressed effectively.

4 The Trial

Funded by the Research Councils UK Digital Economy Programme as part of the CHARM project, the six-week trial of the bActive app was conducted between October and December 2011 with 152 participants from Bristol, UK. Following the principles of a randomised controlled trial, participants were randomly assigned to one of a *control condition* (no feedback and no access to the interactive elements of the app), an *individual condition* (feedback on their own walking only – see Figure 3) and a *social norms condition* (feedback that also included social norms data).



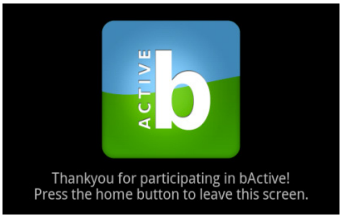


Fig. 3. The initial bActive screen as seen by those in the individual condition (Left) and control condition (Top)

On-street recruitment was conducted by a market research agency in twenty locations spread across Bristol. The incentive for participation was the study phone, the HTC Desire-S, which participants were able to keep. Recruits were told that the purpose of the study was to measure the amount of walking people did. Importantly, and unlike in most previous studies, participants were *not* asked to walk more. Each recruit had to agree to put his current SIM card into the phone, use it as his main mobile and carry it in his trouser pocket for the duration of the study. To ensure that users in the social norms condition were able to compare themselves to a group of broadly similar people, it was important to focus on a specific segment of users. Although the study could have been conducted with either men or women, it was decided to focus on men because of the need for the study phones to be carried in trouser pockets and the likelihood that women would find it more difficult to comply with this requirement. The age-range of 22-40 was chosen in order to avoid the health risks associated with the over-exertion of older users.

Of the 152 participants, 78% were in employment, 14% were students and 8% were unemployed. Of those in employment, 50% were in sedentary jobs (e.g. office workers), 44% were in moderately active jobs (e.g. teachers) and 6% were in very active occupations (e.g. postal workers). Meanwhile, 59% of the participants said that they regularly engaged in sport (about the same as the national average for this demographic [22]) and 63% owned some form of motorized vehicle.

The intention of the research team had been to collect a week of baseline walking data prior to the feedback phase of the trial, but a technical malfunction during this initial week rendered the data unusable, so no baseline comparison was possible.

During the 6-week feedback phase of the trial, participants were sent regular emails and SMSs reminding them to keep the phones in their pockets; with further prompts sent to those whose phones had not sent data for more than a day. In addition, those in the individual and social norms conditions were sent a weekly motivational text message (e.g. "Walking is one of the best activities for your health. How much are you doing? Check the app!")

During the trial, exercise and app use data was collected on a daily basis. Specific variables collected were the number of steps taken by each participant and, for those in the two experimental conditions that were provided with some kind of feedback, the number of times per day they activated the bActive app. In all, the resulting dataset consisted of up to 42 daily observations for each of the 152 respondents across the 6 week period of the trial (6,214 observations in total).

Prior to the trial and at its close, an online questionnaire collected demographic data and potentially confounding variables such as prior use of smartphones, patterns of physical activity, attitudes to physical activity and perceived impacts of the trial.

Subsequent to the trial, two waves of interviews were conducted with trial participants. Sampling criteria for wave-1 (N=7), conducted one to two months after the trial by a member of the research team, were feedback condition and self-reported changes in walking behaviour. For wave-2 (N=8), conducted ten months after the trial, sample selection focussed on those still using the app in April 2012 – four months after the end of the trial – and those with a below average total number of steps. Participants were approached by telephone and offered a £20 incentive.

Table 1. Features of the interview sample

Month of interview	Total	Feedback condition: C- Control I- Individual S- Social			Participant age		Living with children[1]	Self-reported change in walking[2]			Still using app April 2012	Self-reported comparison to average[3] A- above B- below C- varying		
		C	I	S	20s	30s		+	0	-		A	B	C
Jan 2012	7	2	2	3	5	2	3	4	2	1	n/a	1	1	1
Oct 2012	8	0	3	5	5	3	3	7	0	1	3	0	4	1

[1]The number of participants that had children living with them

[2]The number reporting that their walking had increased because of the study (+), stayed unchanged (0) or decreased (-)

[3]The number in the social norms condition reporting that during the study their step-count had usually been above average (A); below average (B), varied between above and below average (C)

The trial set out to test three hypotheses:

- H1 – those with access to feedback will have higher step-counts than those in the control condition
- H2 – those in the social norms condition will have higher step-counts than those in the individual feedback condition
- H3 – those in the social norms condition will use the app more often than those in the individual feedback condition

5 Analysis Method

Given the structure of the data, which had multiple daily observations nested within each participant, longitudinal multilevel modelling [35] was used to test the three hypotheses. As described below, the same 3-stage series of analyses was performed for H3 (where the outcome variable was frequency of app use per day) as was used for H1 and H2 (where the outcome variable was number of steps per day).

First, an unconditional model with no predictors was run in order to calculate the ICC(1) statistic (the percentage of variance in scores over time attributable to differences between participants) and the variance to be explained at the within- and between-participant levels. Second, a fixed growth model was fitted to the data in order to estimate the shape and direction of changes to the outcome over time. To do this, linear and quadratic effects of time (i.e. days since the start of the study) were added to the model, together with a dummy variable for day of the week (with Sunday set as the reference category).

The third stage tested for variability of change between participants by allowing the coefficients of the growth parameters (i.e. the linear and squared effects of time) to become random (i.e. to vary by participant). Potentially confounding participant-level control variables were then added: i.e. marital status; number of children under

17 in the household; employment status (30+ hours' employment per week; 8-29 hours' employment per week; carer/unemployed and in receipt of benefits; student, or self-employed); ownership and use of a motorized vehicle, motorcycle or bicycle (each coded separately), and previous ownership of a smartphone.

Finally, the effects of experimental condition and its interaction with the time-point were added to the model to see if they accounted for between-participant variation in the level of the outcome and its change over time. At each stage, model improvement was evaluated by testing the reduction in the model deviance and assessing the extra variance explained at both within- and between-participant levels.

There were two outcome variables. The first, Steps, was derived from the step-count, which was log-transformed to negate the impact of outliers (participants with extremely high step-counts). The distribution of the second outcome variable, App Use (the number of times per day users activated the bActive app), was severely positively skewed. For this reason, data for the participants who could access the app interface (the two experimental conditions; $N = 110$) were analyzed using a multilevel generalized linear model, treating the error distribution as Poisson and applying a logarithmic link function. Given the large sample size at the level of the time-point, study day ($N = 6214$ for Steps; $N = 4229$ for App Use), a significance level of $p < 0.0005$ was used for assessing the acceptance or rejection of null hypotheses at this level. For participant-level effects ($N = 152$ for Steps; $N = 102$ for App Use) the more typical $p < 0.05$ level of significance was applied. Where hypotheses were directional, one-tailed tests were used.

The interviews and focus groups were transcribed, coded using Atlas-ti and analysed using a combination of thematic and discourse analysis.

6 Results

6.1 Impacts of Feedback on Step-Counts

An assessment of the variation in *Steps* revealed a high level of clustering within participants, with an ICC(1) statistic of 0.33 indicating that a third of the total variation was due to between-participant differences. The introduction of linear and quadratic effects of time alongside dummy codes for day of the week explained a statistically significant but small 4% of within-participant variance and almost no between-participants variance. Tests of fixed effects coefficients indicated that of these three predictors (the linear effect of time, the quadratic (curvilinear) effect of time and day of the week), the third was the primary explanatory variable. Since quadratic change offered no improvement over a simple linear effect, it was dropped from the model.

There was evidence that the linear effect of time varied between individuals. When this random effect was added, along with the covariance between starting level and extent of linear change, the model deviance reduced significantly (Δ Deviance = 111 on 2df, $p < 0.0005$) and the unexplained within-participants variance was reduced by a further 4%. Of the demographic and control variables, only employment status and car ownership had a significant effect upon Steps, with full-time and part-time employees likely to have a higher step-count than other groups and car owners likely to have lower step-counts than non-car owners.

The tests for hypotheses H1 and H2 show that Experimental Condition (a dummy variable coded with control group as the reference category) had a statistically significant effect (Individual vs. Control: B = 0.474, $p < 0.05$; Social norms vs. Control: B = 0.526, $p < 0.05$), explaining a further 7.7% of the between-participants variance in the step-count. Compared to those in the control condition, the average expected step-count of those in the individual feedback condition was 59% higher and that for the social feedback condition was 69% higher (an average of 64% for the two experimental conditions). The null hypothesis for H1 was therefore rejected in favor of the finding that those receiving feedback had higher step-counts than those in the control condition. However, there was no significant difference in Steps between the two experimental conditions, so the null hypothesis for H2 could not be rejected.

There was no evidence that Experimental Condition had any effect on between-participants variation in the rate of change, over time, in the step-count. The interaction of experimental condition and time point was not statistically significant and this interaction reduced the model deviance by just 1 on 1df ($p > 0.05$), explaining only 0.6% of the variation in slopes.

Figure 4 shows the temporal variation in step-counts and illustrates the findings presented above.

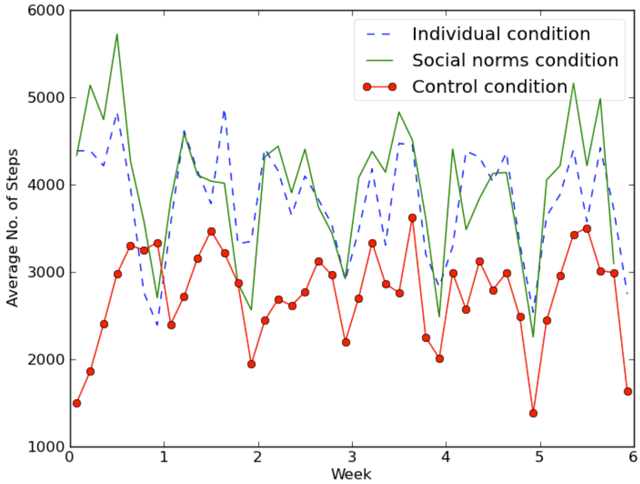


Fig. 4. Daily average step-counts for the three experimental conditions

6.2 Impacts of Type of Feedback on Engagement with the App

Over the course of the study, those in the two feedback conditions opened the app an average 3.9 times per day (median = 3.5, SD = 2.6), on each occasion keeping it open and visible on the screen for an average of 32.0 seconds (median = 33, SD = 9.0). Indeed, 87% of those in the individual condition and 89% of those in the social norms condition used it every day for half or more of the study days, and in the final week of the study participants from the two feedback conditions were still opening the app on average 2.3 times a day (SD = 1.9; median = 1.9). The survey evidence suggests that these figures reflect genuine enthusiasm for the app. Of those in the two feedback conditions 91% reported that the app was 'interesting', 67% that it was 'fun' and 73% that they would continue to use the app after the trial. Furthermore, only 19% reported losing interest in the app before the study end, only 15% of participants from the two feedback conditions reported that the step-count had not been "accurate enough for my needs" and only 11% that "lack of accuracy caused me to use the app less". The absence of any evidence for a non-zero correlation between perceived accuracy and either *Steps* or *App Use* indicates that problems with accuracy had little effect on the impacts of the app on behaviour.

As with Steps, variability in App Use was highly clustered within individuals.

As illustrated in Figure 5, the app was opened most often in the first few days of the study, with usage thereafter declining – first rapidly and then more gently (i.e. a curvilinear effect of study time-point was found to be statistically significant). This decrease over time varied between participants in both shape and rapidity, as evidenced by the fact that the addition of random effects of the time and time-squared terms (and the covariances between intercept and slope) increased the goodness of fit of the model (Δ Deviance = 1288 on 5df, $p < 0.0005$).

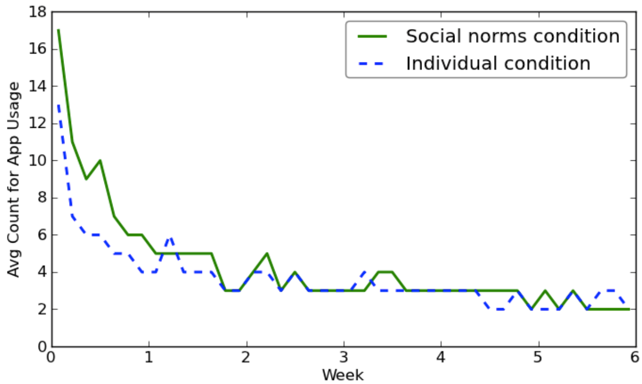


Fig. 5. The average daily frequency with which participants opened the bActive application on their phones

None of the participant-level control variables had a statistically significant impact on *App Use*. This is somewhat surprising, for men with children and those in employment (especially full-time employment) might be expected to have less time for such activities. The lack of any such effect suggests that interest in the feedback was sufficient for participants to use it in spite of other, conflicting, calls on their time and attention. The null hypothesis for H3 could not be rejected, for no difference was found between the two experimental conditions with regard to level or change of use.

6.3 Interview Findings

The interviews suggest that the increase in *Steps* by those in the intervention condition was the result of a number of features of the app design. Two of these were of particular importance: 1) the graphical display of walking patterns in the *Past Wee*k and *History* screens and 2) the always-on design, which meant that the app measured steps continually without having to be switched on. These features made users more aware of incidental walking. Interview respondents reported that before using the app they tended not to be aware of walking that was incidental to the achievement of other activities (e.g. the walking involved in shopping or work). The app made such walking more apparent, prompting one interviewee to comment, "...walking more than I thought, yeah. [...it is] surprising how much walking you do; just little bits here and there, walking around the workshop." This, in turn, showed users that they could become more active without having to engage in entirely new physical activities, and that all they needed to do was include more incidental walking in their existing activities.

The interview data further suggests that simply being seen to measure walking, the app encouraged users to view it as an activity in its own right. Where previously, walking had been "just something you do because it's a natural thing", it now became a measurable exercise that was subject to target-setting. Measurement also encouraged walking by helping users assess what was achievable. Before using the app, some users had relatively little understanding of distances, and the prospect of, for example, a two-mile walk might have been daunting ("Phew! Two miles, I'll hop on a bus" – David). By showing people how many miles they walked in the course of an ordinary day, the app made the concept of 'a mile' more familiar and thus made it more feasible for them to walk a number of miles ("Actually I walked two miles the other day and it seemed like nothing; I can walk that" – David). Finally, measurement highlights the difference between days with lots of incidental walking and days that are less active. This encourages users to walk more so that they are not obliged to see themselves as "lazy" or as "dossing".

The interviews do not fully explain the absence of an incremental impact for the social norms feedback. On the one hand, they suggest that those receiving social feedback became competitive and walked more when they thought they could "win" or "beat the average". On the other hand, there is some evidence that those who were below the average, either or a particular day or more generally, were less likely to walk more because they felt there was no possibility of 'winning'.

7 Discussion and Conclusions

This study indicates that always-on, accelerometer-based smartphone apps can increase walking amongst males by around 64%. This degree of behaviour change could have very real benefits for the general health of users and the prevalence of obesity and chronic disease in the population.

bActive's always-on feature allowed it to measure the walking inherent in practices not initially considered by users as 'walking' or 'exercise' and highlighted the periods in which they were relatively inactive. This had a transformative effect on some users, motivating them to avoid inactivity, making walking more of an activity in its own right, giving users the confidence that they could walk longer distances than they had previously realized and helping them to see that they could increase their activity levels simply by changing the way they conducted their usual activities.

Furthermore, there participants found the app engaging and enjoyable to use; with the trend data in the Last Week and History screens, in particular, holding their interest. Although use of the app fell away quickly from its initial high, the rate of decline slowed rapidly; participants were still accessing the app almost four times a day by the end of the study, and many expressed an interest in continuing to use it.

These results were achieved without the provision of any program of support or instruction, for apart from the feedback displays, the only input received by participants was the weekly motivational text message. This distinguishes the bActive evaluation from many previous studies, including many of those in the meta-analysis of pedometer interventions amongst the general population by Kang et al [23], the largest of which [21,40] included extensive additional programs of motivation or instruction.

The absence of any evidence in support of hypotheses H2 and H3 can be interpreted either as reflecting on the design of the social feedback or as evidence that 'social sharing' is not as important for the promotion of exercise as has previously been argued (e.g. [13]). The evidence in the literature on this issue is surprisingly weak, with assertions sometimes being made with little apparent empirical support. However, the lack of clear evidence for the effectiveness of social sharing does not allow its importance to be dismissed. Users were told that the comparison group comprised other males of about the same age who lived in Bristol, but it is possible that the salience and effectiveness of the comparisons would have been greater had users been given control over the types of people included. In addition, social sharing might have been approached in an entirely different way such as, for example, by allowing users to see the progress of other individual members of the comparison group.

A suggestion for future studies might be to increase the salience of the comparison group by defining it more tightly. Practitioners of the social norms approach generally argue that the most effective reference group comprises those whom participants consider most like themselves – e.g. [4,24]. This might, for example, mean separating those in physically active occupations from those doing more sedentary work.

A final consideration is the design of the feedback display. It is possible that the display of the social norms data on the same graph as the individual feedback detracted from the impact of the latter and that this, and not any lack of impact of the social data itself, weakened the impact of the full version of the app. Given that the individual feedback alone was associated with an increase in walking of 64%, it is clear that this should be an important element of any feedback strategy. However, when the social norm is much higher, even on just one day of the week, than a person's own highest step-count, the curve for the person's own steps becomes flattened and the all-important variations in his own performance are less obvious.

Minor issues with battery power and measurement accuracy notwithstanding, it is clear that when it comes to measuring and encouraging active lifestyles, the ubiquity of smartphones lends phone-based apps an important advantage over dedicated pedometer devices that require up-front commitment to increased exercise. Furthermore, the passive nature of data collection used by the bActive app encouraged use of the app and allowed users to collect data effortlessly and continually.

Those using the app recorded on average 64% more steps than those that did not. This highlights the power of the bActive approach and draws attention to the potential for the use of individual-level feedback that encourages people to reflect on the patterns in their own behaviour, identify opportunities for change and realise those opportunities. From the evidence in this study, it seems likely that an approach modelled on bActive could have a real positive impact on the health and fitness of the population.

Acknowledgements. The authors would like to thank the Research Councils UK Digital Economy Programme for supporting the bActive study and the wider CHARM project of which it was a part. Thanks, too, to Mubaloo, Bristol, for their constructive technical comments on the bActive app.

References

1. Allcott, H.: Social Norms and Energy Conservation. J. Public Econ. 95(9-10), 1082–1095 (2011)
2. Andersen, L.B.: Physical Activity and Health: Even Low Intensity Exercise Such as Walking is Associated with Better Health. Brit. Med. J. 334, 1173 (2007)
3. Baker, G., Mutrie, N., Lowry, R.: Using Pedometers as Motivational Tools: Are Goals Set in Steps More Effective Than Goals Set in Minutes for Increasing Walking? International Journal of Health Promotion and Education 46 (2008)
4. Berkowitz, A.D.: The Social Norms Approach: Theory, Research, and Annotated Bibliography (2004), http://www.alanberkowitz.com
5. Bohannon, R.W.: Number of Pedometer-Assessed Steps Taken Per Day by Adults: A Descriptive Meta-Analysis. Phys. Ther. 87(12), 1642–1650 (2007)
6. Bosari, B., Carey, K.B.: Descriptive and Injunctive Norms in College Drinking: A Meta-analytic Integration. J. Stud. Alcohol. 64(3), 331–341 (2003)
7. Bravata, D.M., Smith-Spangler, C., Sundaram, V., Gienger, A.L., Lin, N., Lewis, R., Stave, C.D., Olkin, I., Sirard, J.R.: Using Pedometers to Increase Physical Activity and Improve Health: A Systematic Review. J. Am. Med. Assoc. 298(19), 2296–2304 (2007)
8. Brusso, R.C., Orvis, K.A.: The Impeding Role of Initial Unrealistic Goal-Setting on Videogame-Based Training Performance: Identifying Underpinning Processes and a Solution. Computers in Human Behavior 29(4), 1686–1694 (2013)
9. Burchell, K., Rettie, R., Patel, K.: Marketing Social Norms: Social Marketing and the 'Social Norm Approach'. Journal of Consumer Behaviour 12, 1–9 (2013)
10. Choi, B.C.K., Pak, A.W.P., Choi, J.C.L., Choi, E.C.L.: Daily Step Goal of 10,000 Steps: A Literature Review. Clinical Investigative Medicine 30(3), E146–E151 (2007)
11. Cialdini, R.B., Goldstein, N.J.: Social Influence: Compliance and Conformity. Annu. Rev. Psychol. 55, 591–621 (2004)
12. Cialdini, R.B., Reno, R.R., Kallgren, C.A.: A Focus Theory of Normative Conduct: Recycling the Concept of Norms to Reduce Littering in Public Places. J. Pers. Soc. Psychol. 58(6), 1015–1026 (1990)

13. Consolvo, S., Everitt, K., Smith, I., Landay, J.A.: Design Requirements for Technologies that Encourage Physical Activity. In: CHI 2006, pp. 457–466. ACM (2006)
14. Consolvo, S., Klasnja, P., McDonald, D.W., Avrahami, D., Froehlich, J., LeGrand, L., Libby, R., Mosher, K., Landay, J.A.: Flowers or a Robot Army? Encouraging Awareness & Activity with Personal, Mobile Displays. In: UBICOMP 10, pp. 54–63. ACM (2008)
15. Consolvo, S., McDonald, D.W., Toscos, T., Chen, M.Y., Froehlich, J., Harrison, B., Klasnja, P., LaMarca, A., LeGrand, L., Libby, R., Smith, I., Landay, J.A.: Activity Sensing in the Wild: A Field Trial of Ubifit Garden. In: CHI 2008, pp. 1797–1806. ACM (2008)
16. Dewa, C.S., deRuiter, W., Chau, N., Karioja, K.: Walking for Wellness: Using Pedometers to Decrease Sedentary Behaviour and Promote Mental Health. The International Journal of Mental Health Promotion 11(2), 24–28 (2009)
17. Groves, A.: Using Pedometers as a Valid Method of Determining Physical Activity Intensity Level. MSc thesis, Brigham Young University, Dept. of Exercise Sciences (2008)
18. Harries, T., Rettie, R., Studley, M., Chambers, S., Burchell, K.: Is social norms marketing effective? A case study in domestic electricity consumption. European Journal of Marketing (in press)
19. Ilies, R., Judge, T.: Goal Regulation Across Time: The Effects of Feedback and Affect. J. Appl. Psychol. 90(3), 453–467 (2005)
20. Iwane, M., Arita, M., Tomimoto, S.: Walking 10,000 Steps/Day or More Reduces Blood Pressure and Sympathetic Nerve Activity in Mild Essential Hypertension. Hypertens. Res. 23, 573–580 (2000)
21. Jackson, E.M., Howton, A.: Increasing Walking in College Students Using a Pedometer Intervention: Differences According to Body Mass Index. J. Am. Coll. Health 57(2), 159–164 (2008)
22. Jones, H., Millward, P., Buraimo, B.: Analysis of the Taking Part Survey. University of Lancashire, Lancaster (2011)
23. Kang, M., Marshall, S.J., Barreira, T.V., Lee, J.-O.: Effect of Pedometer-Based Physical Activity Interventions: A Meta-Analysis. Res. Q. Exercise Sport 80(3), 648–655 (2009)
24. Lewis, M.A., Neighbors, C.: Gender-Specific Misperceptions of College Student Drinking Norms. Psychol. Addict. Behav. 18, 334–339 (2004)
25. Lin, J.J., Mamykina, L., Lindtner, S., Delajoux, G., Strub, H.B.: Fish'n'Steps: Encouraging Physical Activity with an Interactive Computer Game. In: Dourish, P., Friday, A. (eds.) UbiComp 2006. LNCS, vol. 4206, pp. 261–278. Springer, Heidelberg (2006)
26. Locke, E., Latham, G.: Building a Practically Useful Theory of Goal Setting and Task Motivation: A 35-Year Odyssey. Am. Psychol. 57(9), 705–717 (2002)
27. Locke, E., Latham, G.: New Directions in Goal-Setting Theory. Curr. Dir. Psychol. Sci. 15(5), 265–268 (2006)
28. MacLellan, G., Baillie, L., Granat, M.: The Application of a Physical Activity and Location Measurement System to Public Health Interventions to Promote Physical Activity. In: Proceedings of the 2nd International Conference on Pervasive Technologies Related to Assistive Environments, pp. 1–8. ACM, New York (2009)
29. McCormack, G., Milligan, R., Giles-Corti, B., Clarkson, J.P.: Physical Activity Levels of Western Australian Adults 2002: Results from the Adult Physical Activity Survey and Pedometer Study. Western Australian Government, Perth (2003)
30. Mladenov, M., Mock, M.: A Step Counter Service for Java-Enabled Devices Using a Built-in Accelerometer. In: Proceedings of the Fourth International Conference on Communication System Software and Middleware, pp. 1–5. ACM (2009)
31. Morris, J., Hardman, A.: Walking to health. Sports Med. 23(5), 306–332 (1997)

32. Pooley, C., Tight, M., Jones, T., Horton, D., Scheldeman, G., Jopson, A., Mullen, C., Chisholm, A., Strano, E., Constantine, S.: Understanding Walking and Cycling: Summary of Key Findings and Recommendations. University of Lancaster, Lancaster (2011)
33. Schneider, P.L., Bassett, D.R., Thompson, D.L.: Effects of 10,000 Steps per Day Goal in Overweight Adults. Am. J. Health Promot. 21(2), 85–89 (2006)
34. Schultz, P.W., Nolan, J.M., Cialdini, R.B., Goldstein, N.J., Griskevicius, V.: The Constructive, Destructive, and Reconstructive Power of Social Norms. Psychol. Sci. 18(5), 429–434 (2007)
35. Singer, J.D., Willet, J.B.: Applied Longitudinal Data Analysis – modelling change and event occurrence. Oxford University Press (2003)
36. Steel, P., König, C.J.: Integrating Theories of Motivation. Acad. Manage. Rev. 31, 889–913 (2006)
37. Thaler, R., Sunstein, C.: Nudge: Improving Decisions about Health, Wealth, and Happiness. Yale University Press, New Haven (2008)
38. Toscos, T., Faber, A., An, S., Gandhi, M.P.: Chick Clique: Persuasive Technology to Motivate Teenage Girls to Exercise. Extended Abstracts on Human Factors in Computing Systems, pp. 1873–1878. ACM (2006)
39. Tudor-Locke, C., Ainsworth, B.E., Thompson, R.W., Matthews, C.E.: Comparison of Pedometer and Accelerometer Measures of Free-Living Physical Activity. Med. Sci. Sport. Exer. 34, 2045–2051 (2002)
40. Winnet, R.: Guide to Health: Nutrition and Physical Activity Outcomes of a Group-Randomized Trial of an Internet-Based Intervention in Churches. Ann. Beh. Med. 33(3), 251–261 (2007)

F-Formations in Cooking Together: A Digital Ethnography Using YouTube

Jeni Paay[1], Jesper Kjeldskov[1], Mikael B. Skov[1], and Kenton O'Hara[2]

[1] Aalborg University, 9220 Aalborg, Denmark
[2] Microsoft Research Cambridge CB3 0FB, UK
{jeni,jesper,dubois}@cs.aau.dk, kenton@gwork.org

Abstract. Cooking together is an important part of our lives. We cook with others not only to create a meal, but also to enhance our relationships. But how does this role of communal cooking translate into modern society where families and friends are increasingly separated physically and connected primarily online? Motivated by this question we have embraced research into the design of future networked cooking spaces. The first step has been to understand how people use physical space while cooking together. Through a digital ethnography on YouTube videos, we have analyzed the spatial configurations of people, food and technology based on Kendon's notions of spacing and orientation. Our main contribution is the identification of known F-formations as well as new formations taking place during social cooking. Based on this we suggest that given the presence of formations in the kitchen different from those found during activities that are mainly conversational, simply installing traditional video-conferencing systems in people's kitchens will not suffice in facilitating the interactions taking place there. Instead, designers need to rethink the positioning and use of cameras and displays.

Keywords: Cooking, F-formation, proxemics, digital ethnography, YouTube.

1 Introduction

Recently, the kitchen has become a focus for HCI research. Cooking together is an important part of our lives. We cook with friends, families, colleagues and strangers to share not only the experience of creating a meal, but in the spirit of commensality, as a social event where we enhance our relationships with others through shared stories, relating daily happenings and discovering new ideas about food preparation from each other. Given this role of food and the kitchen as a place for social interaction, it is important to understand the role computer technology might play in this context – now and in the future [8]. But how does this role of communal cooking translate into modern society where families and friends are increasingly separated physically and connected primarily online?

Motivated by this question, we have embraced research into on future cooking spaces that enable people to have meaningful and positive social food experiences

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 37–54, 2013.

while cooking *together*, but in *different places*. As with environments for the work domain [29], future kitchen environments may integrate designs that provide geographically distributed people with the experience of cooking with remote family and friends using their respective kitchens as one shared digitally "blended" cooking space. This, we believe, would give people the opportunity to interact socially with close ones living in another place as an integrated part of their daily lives and household routines, which are important aspects of keeping distant relatives and friends connected [26]. Fundamental to this, we think it is important that people sharing a blended cooking space experience the same essential qualities that make collocated human-food interaction meaningful.

Subscribing, as we do, to the view that kitchens are "sites where meaning is produced, as well as meals" [1] has some profound consequences for the way one does HCI research in this domain. Firstly, it becomes more about people and their interactions in and around the kitchen, rather than just about building new technology to assist kitchen tasks. Secondly, if we wish to design compelling and meaningful new technologies that might enrich these interactions, we need to study people's practices in the kitchen and unravel and understand the richness of social interactions and experiences that happen there while cooking together.

In this paper, we explore how people arrange themselves spatially while cooking together with a "remote viewer" sharing the experience, and how new methodologies may be devised for studying this empirically. Specifically, we seek to better understand the activity of shared cooking: people's interactions with co-present others, with the physical kitchen layout, with cooking artifacts, and with remote viewers watching on video. As a source of data for this study we have turned to the vast amount of user-generated digital video material available on YouTube depicting people cooking together in private kitchens and wanting to share this social experience with an outside audience through the Internet. Analyzing this video material can be viewed as a form of "digital ethnography" [22] where the phenomena of interest are not observed first-hand but are mediated through digital data. This approach, we believe, has great potential for informing interaction design by giving us insight into human behavior and activity in real world settings, such as private and domestic life, that are inherently difficult to reach with traditional ethnographic techniques.

The paper is structured as follows. First we provide a background of related work in two areas: HCI in the kitchen, and analysis of human spatial configurations. We then turn our attention to the specific digital ethnography of cooking together that we have carried out, describing its methodological motivations and challenges and the details of the observational and analytical procedures followed. This is followed by a detailed account of our findings, describing and illustrating the formations identified in the videos, and the activities observed for creating and sustaining them. In the discussion section we take a step back from our empirical findings and reflect on their significance and implications. As the methodology used in our study is also of some novelty to the CHI community we also reflect on this approach to studying phenomena in domestic settings for the purpose of interaction design. Finally, we conclude on our work and outline plans for further work towards informing design.

2 Background and Related Work

Our research draws on two areas of related work: 1) HCI in the kitchen, and 2) theories about how people organize themselves in physical space.

2.1 HCI in the Kitchen

A number of research efforts have begun exploring digital technologies in the support of cooking. These have included *Cooks Collage* [35], *CookingNavi* [10], *Cooking with the Elements* [4], *The Living Cookbook* [34], *Not Enough Cooks in the Kitchen* [33], *CounterActive* [17], *Kitchen Album* [20], and *PersonalChef* [23]. These technologies vary in terms of how they approach cooking and requirements. For some, the approach has been to support the functional aspects of cooking and overcoming various difficulties with it (e.g. [35] and [10]). For others attention has been on the social meanings and practices of cooking.

According to Grimes and Harper [8], research on HCI in the context of kitchens has for the most part neglected technology design that focuses on the social experiences that we have with each other in the context of cooking and eating. Instead, the overwhelming trend has been to design "corrective technology" applications that act as a remedy for perceived problems that people have with planning, preparing and consuming food [8]. As an alternative, Grimes and Harper advocate a focus on what they call "celebratory technology" – technology that gives us positive, delightful, pleasurable and exciting experiences with food – as a potential new research and design space in HCI. Likewise, in their kitchen manifesto, Bell and Kaye [1] advocate a kitchen/technology relationship that draws on and learns from the rich cultural history of the kitchen as a place for living and above all, a focus on those who experience the space, rather than the resident technologies. The manifesto calls for a focus on experience over efficiency and an understanding of the use of objects in the kitchen context.

Exemplifying this line of thought, research on HCI in the kitchen has recently and increasingly turned its attention to the social meanings and practices of cooking in which the kitchens are viewed as "sites where meaning is produced, as well as meals" [1]. For example, sharing recipes as a social act (e.g. [34]) or providing mentoring support for friends and family while they learn to cook (e.g. [33]).

In *The Living Cookbook* [34], people's cooking experiences are recorded and shared in an interactive digital medium containing video and annotated audio that others may play back later. The agenda is to enable people to share cooking experiences, to educate others in cooking practice, and to suggest a sense of presence and sociability in the cooking activity. The role of technology in this design is to support intimacy, communication, education, fun and creativity in the cooking experience. It also allows people to strengthen cultural and social connections with others, and to stimulate cross-cultural as well as cross-generational interaction.

In *Not Enough Cooks in the Kitchen* [33], video conferencing and video archiving are used to facilitate collaborative cooking experiences, and to receive, record and keep cooking tips from remote friends and family. The main purpose is the teaching and learning of cooking skills by providing a medium for experienced cooks to share their knowledge with novices and giving them valuable life skills and digital keepsakes of the experience. Designed to replace phone calls and emails between family and friends when trying to learn how to cook, it caters for much more of the richness of collocated cooking, and makes cooking with remote partners a socially enjoyable experience.

Similar to these examples, our focus here is concerned with the shared experience of cooking, and using technology for bringing people across different households together socially in shared spaces for cooking. But in order to meaningfully inform technology design that are both useful and sensitive to the social and physical context that they are to operate in, we need to understand better people's daily practices in these spaces [1].

2.2 Proxemics and F-Formations

As discussed by Rogers [32], a number of theories from other disciplines, new or relatively unknown to the field of HCI, are currently making their way into our field. Some of these, she argues, represent "the beginnings of a 'turn to space'", including schematic frameworks for modeling people's interactions with each other in physical space, such as *proxemics* and *F-formations.* Developed for analyzing social interactions outside the technology domain, these theories and frameworks have recently proved to be highly valuable in relation to interaction design as this domain broadens itself out to embrace the challenges of digital ecologies and pervasive and ubiquitous computing environments.

Proxemics covers a topic concerned with peoples' use of their immediate physical space in their interaction and communication with each other. It was originally developed by Edward Hall, a cultural anthropologist, in the early 1960s [9], but has recently gained renewed momentum in HCI as a framework for also describing and understanding people's interaction with devices and displays around them. In brief, Hall proposed that (in the USA) four interpersonal distances govern social interaction: *intimate* (0-0.46m), *personal* (0.46-1.22m), *social* (1.22-3.66m) and *public* (>3.66m). In the context of HCI this provides a useful framework for analyzing and designing interactive technology. Examples of this include Obata and Sasaki' virtual visiting system [28] and Heaton's case studies of social interaction in virtual workspaces [13] in a CSCW context, and more recently in a more general interaction design context, the works of Greenberg and colleagues on Proxemic Interactions [7].

F-formations is a part of a conceptual framework about people's interpersonal spacing and orientation developed by Adam Kendon, a leading researcher in the study of gesture, in the late 1970s [18] [19]. The notion of F-formations helps explain how people arrange themselves spatially with respect to others, and to artifacts in different kinds of social interactions and collaborations and can also be used to explore the influence of physical environment on the interpersonal interactions taking place there.

It was developed by Kendon in response to previous studies in proxemics in order to draw attention to the fact that *"participants in occasions of interactions of all sorts may enter together into a cooperative relationship to sustain, through time, a more or less constant pattern of spatial orientational arrangement"* [Kendon 2012, personal communication]. Briefly described, in an F-formation system individuals have a spaced called a *transactional segment,* which is the space where they focus attention and manipulate artifacts. This space is defined in relationship to their lower body, and changes in size depending on the kind of activity people are doing. An F-formation is formed when the transactional segments of two or more people overlap and create a shared inner space, where the shared activity of those people occurs. This is called the *o-space* [18] or the *use-space* [19]. In the latter Kendon gives a series of concrete examples of different types of F-formations in interpersonal communication, such as the L-shaped formation of people standing orthogonal to each other, the vis-à-vis formation of facing each other directly, the side-by-side formations of facing the same direction, and the semi-circular formation of more than two people side-by side curving inwards to speak to each other.

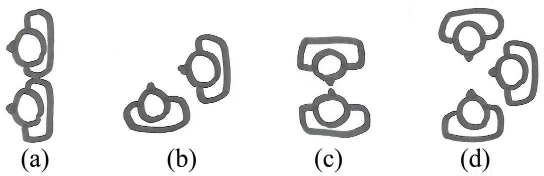


Fig. 1. Schematic examples of different F-formations: (a) side-by-side, (b) L-shaped, (c) vis-á-vis, (d) semi-circular

Like proxemics, the work on F-formations has recently attracted interest among HCI researchers concerned with design and use of technology embedded in the physical environment, for example in [14] [15] [21] [24] [25] [29]. Marshall et al. [21] use the F-formation system for analyzing social interactions between visitors and staff in an information center for tourists. They show how the presence or absence of F-formations can explain the relationship between co-located interactions and the physical environment, and how this may inform technology design. Mentis et al. [24] conduct a similar analysis but in the context of work practices in neurosurgery theatres. This leads to insight about the spatial organization of shared work practices in this specific setting, and to understanding design implications for touchless interaction with medical technology. Other uses of F-formations in HCI include investigating the spatial relationships between humans and service robots during cooperative tasks [15], and detecting F-formations from video footage for analysing social interactive behaviour in crowded environments [14].

Spatial Organization in the Kitchen. Our concerns with the spatial organisation of behavior in the context of shared cooking lies in the fact that architectural design of kitchen spaces, like other spaces [21], structure the kinds of interactions and social activities that are enacted within them. People also adapt their activity patterns in respect to artifacts and others they are interacting with, and the very activity of cooking influences the ways in which human interaction is spatially organized in the kitchen. Working side by side at a kitchen bench, for example, influences how people communicate as opposed to working at opposite sides of a kitchen island. Facing a video camera during the interaction adds yet another level of complexity, not only to local communication, but also to the viewers' perceived distance from the cook, which affects the intimacy of the interaction. Issues such as eye gaze and field of view can all affect the nature of the interaction. How these physical aspects contribute to shaping the experience people have when cooking together is what we have aimed to study and understand.

3 A Digital Ethnography of Cooking Together

Our focus in this paper is on people cooking together and particularly on how they arrange themselves spatially while cooking together. As other studies have done, our approach draws on data from YouTube (e.g. [2] [3] [12] [31]) using it both as a source for video and as "a community that serves as a platform for communication and interaction" [31].

3.1 Using YouTube as Data Source

An important task for HCI researchers interested in the design of new domestic technologies is to observe and make sense of people's daily practices, so that these can meaningfully inform design and seed innovation [1]. While this may sound simple, it actually presents a significant challenge within the context of people's homes, including their kitchens. As the movie "Kitchen Stories" [11] so charmingly illustrates, it is not always ideal, or even desirable, to do traditional direct observation in people's homes. This is also confirmed by research, like Crabtree and Rodden [5], concluding that although the home is relatively easy to access, such direct observation inevitably disrupts the ordinary flow of household activities and can cause people to alter their ordinary behavior.

In response to these challenges, research into the domestic domain has to a large extent turned to methods of self-reporting for obtaining insight into people's daily lives with minimal interference. This has been done through various techniques and data collection media, such as cultural probes (e.g. [6]), technology probes (e.g. [16]), and diaries (e.g. [27]). As an alternative approach, using YouTube as a data source represents a form of "digital ethnography" [22]. In digital ethnography, researchers immerse themselves in the digital world rather than physical and make use of digital data, such as written words, images, audio files, video and online communities, about the phenomenon of interest, which is often available in vast amounts today.

If chosen carefully, Blythe and Cairns argue that YouTube potentially provides several interesting avenues of inquiry as a digital on-line data source [2]. Using YouTube for studying cooking together potentially gives us insight on human behavior and activity in real world settings that are inherently difficult to study using traditional ethnographical technique. As people voluntarily and enthusiastically share a variety of different cooking situations with others on YouTube, we chose this as our data source, with "purposeful sampling" [30, p. 230], for a digital ethnography of how people cook together and share this activity digitally with others. In this way, we were able to rapidly gain access to a breadth of information about a variety of people in different generations, cultures and situations documenting their own experiences of "cooking together" (figure 2).

Fig. 2. YouTube video of "cooking together"

In studying the YouTube videos of people cooking together, our focus was on understanding the social aspects of cooking. In particular, we were interested in how people coordinate the activity, how they organize themselves spatially in respect to physical space and each other, the role of the food in this shared experience, and in particular how they include (or don't include) the viewer of the video in the interaction.

3.2 Analysis

Using the keyword phrase "cooking together", 169 videos were found on YouTube through a search on 15th November 2010. After discarding duplicates and unrelated videos (e.g. a song with the words "cooking together" in the lyrics), a final set of 61 videos of people cooking together remained. Inspired by Blythe and Cairns [3] [3], we conducted a qualitative content analysis of the 61 videos. This resulted in inductive development of the following categories: *family life, family cooking, celebrities cooking, amateur cooking show, professional cooking show, documentary, educational* and *advertising*.

Family life videos typically show people cooking together either in their own homes or outdoors with family or friends. The focus is on the people, their conversations and interactions – showing everyday life in the context of food preparation. There is little or no emphasis on what is being cooked. *Family cooking* videos have family members either taking on the roles of co-cooks, or parents and grandparents involving children in the cooking activity and teaching them how to participate. They are typically located in the family kitchen and sharing cooking techniques with each other is important but actual recipes are seldom elaborated to the viewer. *Celebrities cooking* differs from a cooking show, in that the celebrities are giving the viewer a "snapshot of their ordinary lives" in the context of cooking. These videos often include a background story about the celebrities public life interwoven with the activity of creating food. *Amateur cooking show* videos have the format of a professional television cooking show, with introduction, recipes, techniques and helpful hints, but are created by ordinary people from their home kitchens. They appear to be created specifically for YouTube where production company approval is not necessary to become the host of your own show. *Professional cooking show* videos are often versions of a show that runs on an established television or cable network. These shows have professional chefs and cooking celebrities, typically filmed in a television studio kitchen. *Documentary* videos are similar in some ways to the family life videos, in that there is story unfolding within the food context but in the documentry, people are sharing experiences or historical events with the viewing audience while cooking. *Educational* videos are usually made by a specific educational institution for disemination of cooking skills, to specific learning groups. *Advertising* videos, although often showing people cooking together, do so with the intention of attracting customers to cooking schools, public cooking events and hotels with special cooking weekends.

Secondly, the 61 videos were analyzed from the perspective of Kendon's F-formation system [18] [19]. In this analysis we looked for the following F-formations (for two people) *L-shaped* (standing orthogonal), *vis-à-vis* (facing) and *side-by-side*; and (for groups of three or more people) *circular*, *rectangular*, *semi-circular* and *linear* arrangements. We also identified spatial patterns of interaction between people, artifacts, spaces and the camera (viewer perspective) to give us insight into how people were sharing activities and interacting with others and artifacts in respect to the kitchen's physical layout. For this, we used Kendon's diagramming practice for recording F-formations [18] to "transcribe" the videos, recording patterns of behavior by creating a series of maps (figure 3) of the physical kitchen layout, showing the arrangements and movements of people and cameras within that space.

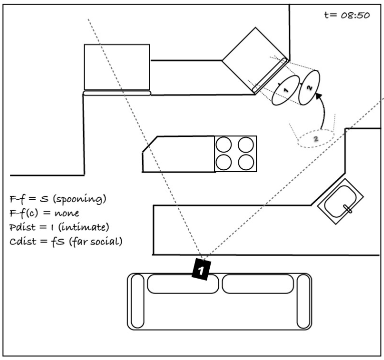


Fig. 3. F-formation Map of YouTube Video 57

Each map is time stamped (*t*), representing elapsed video time, and records a newly established position of individuals. People are shown as numbered ovals with two lines extending to show their transactional segment, and hence where they intersect to create an F-formation. Previous positions (dotted) and paths of movement through spaces are represented, as are artifacts currently being used. We also documented the virtual position of the camera (black square) and the field of view (dotted lines) from that position. This made it possible to identify the viewer's participation in any F-formations. Each person's current activity, the focus of the current camera shot, and the general field of view as a description (e.g., half of kitchen bench top in view) were also noted. These maps were then analyzed to identify the following: the F-formation between co-located cooks (*F-f*); the F-formation between us (the viewer) and the cooks (*F-f(c)*); the distance between cooks (*Pdist*); and the perceived distance between us (the viewer) and the cook *(Cdist)* using distance classifications of Hall [9]. For example, in figure 3, *F-f=spooning*, *Pdist=intimate*. This coding was then studied to identify spatial patterns supporting social connections.

4 Findings

In the following, we outline key findings from our analysis based on the spatial patterns unfolded in the videos. We identified patterns that aligned with the F-formations as described and illustrated by Kendon [19]. In addition to this, we also identified a number of new patterns not described in previous work.

In our analysis we identified a total of 1592 scenes in the 61 videos (a scene being defined as a move to a new person formation or a new camera position). No discernable formation was found in 663 scenes. The remaining 929 scenes involved two or more cooks in different formations and were analyzed and coded in detail.

4.1 Known F-Formations

In the analysis we identified patterns of interaction between cooks that aligned with Kendon's F-formations and were noticeably important for the shared cooking interaction. We identified the following four formations: *Side-by-side, L-shaped*, *vis-à-vis*, and *semi-circular*. Not surprisingly, we found no circular or rectangular formations as they would normally only occur for free-standing groups [21]. The following are examples of these four formations as they were instantiated in the activity of cooking together in their own kitchens, drawn from the sample YouTube videos.

The L-Shaped Formation. Kendon describes the L-shaped formation as participants standing orthogonal to each other [19]. This formation was also extensively used in the videos and was observed in 235 scenes across 43 videos. The L-shaped formation typically forms at the corner of a bench, table or stove-top, where cooks are positioned on two connected sides facing toward a central cooking task on which they are focused. Because this particular formation makes a very intimate shared central activity space, it often facilitates cooks working together on the same cooking task. One cook may be demonstrating a technique, either food preparation at a bench top or stirring food in a pot at a cook top, while the others watch.

Alternatively, it is a good configuration for passing a task backwards and forwards between participants. In the situation of working with small children, the child can

Fig. 4. A YouTube scene illustrating a L-shaped formation with three cooks

also be placed sitting on the bench top, so that they can see clearly and participate equally from that position in the food preparation (see figure 4).

The Side-by-Side Formation. The side-by-side formation occurred in 218 scenes. It was the single most used formation, and was observed in 46 videos. Side-by-side formation occurred in the kitchen when two or more cooks worked along one side of a shared bench or stove top. They may be working on a shared task, but more usually we observed that the cooks in side-by-side formation were working on concurrent but independent tasks (figure 5, bottom left). Another situation when this occurs is during instruction where one or more of the participants may be watching the main cook demonstrate a technique (figure 5, top left).

Side-by-side **Semi-circular** **Vis-à-vis**

Fig. 5. Different YouTube scenes illustrating three F-formations

With the added dimension of the camera in the interaction, the cooks may just be standing or sitting side be side and addressing the camera directly, hence facing the same direction while having a conversation and including the viewer. The use of side-by-side is not surprising as this formation illustrates people "that are jointly concerned about something in their immediate environment" [19]. In the kitchen, food acts as this "something" in the immediate environment and the kitchen layout (e.g. benches, table-tops) would sometimes support or enforce this formation.

The Vis-à-Vis Formation. Vis-à-vis formations (where people arrange themselves face to face) were less common than the two previously illustrated formations. We identified vis-à-vis formations in 63 scenes from 25 videos. The layout of the kitchen certainly had an impact on how the vis-à-vis formation presented itself. The kitchen island design or a centrally placed kitchen table or cook-top with opposite sides accessible for working encourages this formation in food preparation. Cooks stand opposite each other while working on their cooking tasks and converse with each other face-to-face. In the scene in figure 5 (top middle), a father and his daughter work cooperatively on a shared task of shaping zucchini fritters. In the kitchen designs we observed, the width of the table or bench mostly precluded working on a shared task, but cooks could work cooperatively, in that one might be reading the recipe, while the other carries out the instructions (figure 5, bottom middle).

The Semi-Circular Formation. The semi-circular formation only happens in groups of more than two. In our sample videos, we identified 65 scenes from 16 videos involving three cooks. For these three cooks to form a semi-circle, they would generally be gathered around a shared task being carried out in a central shared space. The configuration is semi-circular in shape, that is, one side of the circle is open. This open side in the cooking together context was sometimes a consequence of the kitchen layout. This happened when people gathered around a stove or when younger children were placed sitting on the bench top arranged around the food preparation area (figure 5, top right). In other instances, it appears that people in the kitchen have left one side of the circle open to give the camera a view of their activity (figure 5, bottom right).

4.2 New Formations

While we identified 581 spatial-orientational arrangements that could be described using the known formations from Kendon's work [18] [19], we also found 348 scenes that did not fit with this system. By analyzing these 348 scenes of different spatial-orientational arrangements further, we have identified and named four additional types of formations according to their conversational characteristics. We have named these: *wide V-shaped*, *Spooning*, *Z-shaped*, and *reverse L-shaped* (figure 6).

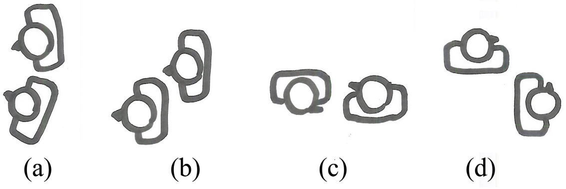


Fig. 6. Schematic examples of new formations:
(a) wide V-shaped, (b) Spooning, (c) Z-shaped, (d) reverse L-shaped

The following are examples of these four formations as they were instantiated in the activity of cooking together.

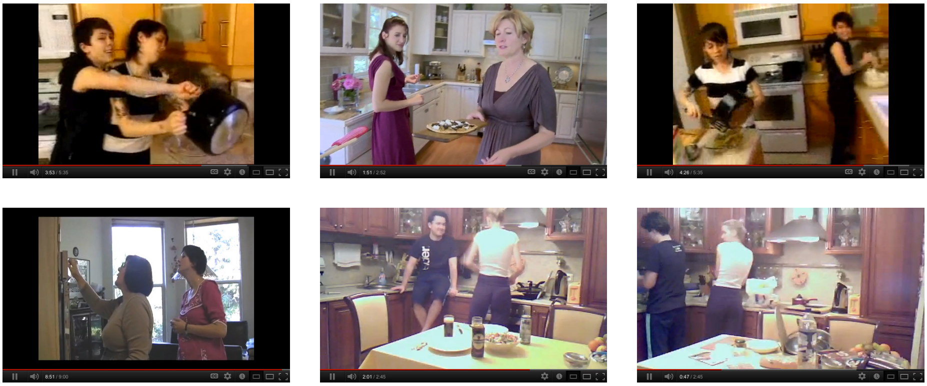
The Wide V-Shaped Formation. The wide V-shaped formation is something that is peculiar to cooks positioned at a long bench top, where they are conversing or working closely together on a single task. We identified this formation in 163 scenes in 38 videos. In this formation, rather than having a side-by-side configuration, where they face forward, working in their own section of bench, they tilt their bodies slightly so that they are facing each other a bit more, but still actively engaging with an activity on the bench, or in the shared space between them. As an example, the two participants in figure 8 are having a conversation and the physical layout of the bench enforces this formation. Another example of a wide V-shaped formation was found when a father illustrates and instructs a child on how to cook.

We further noticed that if the wide V-shape becomes too acute then the participants lose connection with the bench and their cooking activity. There is a continuum of angles between the side-by-side and the L-shape formations but because of the way people work in kitchens and the popular and functional kitchen design of long bench spaces, cooking scenes often involved this very specific formation.

Fig. 7. A scene illustrating a wide V-shaped formation

The Spooning Formation. The spooning formation describes an intimate interaction important in showing and sharing progress while cooking together. We identified 80 spooning scenes in 31 of the 61 videos. When "spooning" in a kitchen, one cook would typically approach another cook from behind to assist and to see what they are doing. This could be in front of a stovetop, or kitchen bench. This means that they are facing the same direction and share the same immediate space in front of them both. From this position they can assist each other in a difficult task (figure 8, top left) or show details of their current activity, such as setting the oven temperature, to each other (figure 8, bottom left). Sometimes this happens because there is only a small amount of space in which to show the second cook what the first is doing, such as when the first cook is working in a corner of the kitchen, or at a narrow sink.

The Z-Shaped Formation. The Z-shaped formation describes the interaction between two people cooking together while standing side-by-side but facing opposite directions. This formation was found in 57 scenes from 21 of the videos. Interaction takes place either by talking away from the other person, each person facing and talking forward, or by turning one's head towards the other while the torso and hands remain directed towards "ones' own work space". This is a configuration that would seldom happen in a conversational context, but is quite a comfortable way of interacting in the cooking together context. For example, a participant sitting on the bench top while talking to or interacting with the cook who is working at the same bench top (figure 8, bottom middle). Further, we saw this formation when cooks pass each other in the constrained environment of the kitchen, while keeping each other updated on current activity or simply continuing their conversation while moving between the variety of tasks and artifacts involved in preparing food (figure 8, top middle).



Spooning **Z-shaped** **Reverse L-shaped**

Fig. 8. Different YouTube scenes illustrating new formations

The Reverse L-Shaped Formation. The reverse L-shaped configuration is another configuration peculiar to the cooking environment. In this configuration, two cooks have their bodies in an L-shaped configuration, but they are facing away from each other. We identified this formation 48 times in 21 videos. In an L-shaped formation, cooks are evidently working on different tasks, but they are coordinating, cooperating and conversing as an integral part of the production of food. Kitchen layout plays a significant role in facilitating this formation, because it is sustained by working at a bench top, stove top or sink that places the two cooks in this position. Figure 8 (right) illustrates two different kitchens and two different examples of reverse L-shaped formations. In the bottom picture, the two cooks are standing in the corner of the kitchen, enforcing an L-shape, but in reverse. As it is quite usual in kitchen design to have benches along the walls of a square room, there are inevitably corners in which two bench-tops meet. Here, two cooks can be working at right angles to each other, while talking over their shoulders to each other, or leaning over to check each other's progress (figure 8, bottom right). Alternatively a central kitchen island makes it possible for one cook to work there, while the other works at a orthogonal bench behind them (figure 8, top right). Again the two cooks continue to converse and cooperate from this position periodically rotating their heads to visually engage with one another.

5 Discussion and Conclusions

In this paper we have explored the experience of how people cook together with a particular emphasis on how they share these experiences through the use of a video camera. More specifically we were concerned with the ways that people oriented towards each other and a range of cooking artefacts in the context of these shared experiences and how such configurations were organized with respect to particular architectural features of the kitchen space. In adopting such analytic concerns, our aim was to provide a resource for thinking about how such activities might be enabled in distributed contexts. That is, by understanding some of the ways that people configure themselves both with respect to features of the local environment and other remote actors and audiences, how might we exploit these features in the configuration of distributed cooking experiences.

One of the ways that we have chosen to explore these concerns, and that forms the particular focus of the current paper, is through an analysis of a rather unusual data resource. This data resource consisted of corpus of existing YouTube videos created and posted by people documenting their various forms of shared cooking experiences. What is intriguing about such a resource is that it provides a view into a wide range of kitchen environments and cooking experiences, revealing both insights into local configuration patterns, but also some particular ways that people choose to organize and present themselves and their environment to a remote audience as mediated by a video camera. While the primary focus of the paper is with what is revealed about the shared cooking experience under such video mediated circumstances, a secondary concern is with some reflections on this unusual methodological approach.

5.1 Spatial Arrangements in Cooking Together

Drawing inspiration from Kendon's F-formation framework for understanding spatial configurations in social interactions, what was apparent from our video analysis was that there were a wide range of spatial configurations adopted by people in the context of sharing their cooking experiences. While some of these were of the vis-à-vis variety that one might typically associate with a standard videoconference model, what was of significance was the number of other configurations identified. Of these, we observed the use of three configurations that aligned closely with those identified in Kendon's original set of conversational arrangements, namely side-by-side, L-shaped, and semi-circular. But the analysis also revealed the use of four additional patterns at play in these contexts, namely wide V-shape, spooning, Z-shape and reverse L-shape.

While the successful adoption of these varied different spatial configurations is of interest in itself, what is of particular concern is why such patterns are in evidence in these cooking experiences. That is, what is it about the activities that constitute these experiences and what is it about the environmental contexts that elicit these patterns? Firstly, kitchens are not free-form neutral spaces in which the collaborators can configure themselves without constraint. Rather, we see how people configure themselves with respect to particular physical features of the kitchen layout (e.g. benches, stoves, tables) as they engage in various cooking activities. As an example, several cooks from our videos formed L-shaped formations at the corner of a bench or stovetop where they would be positioned on two connected sides facing toward a central shared cooking task. It was the physical kitchen layout that encouraged the cooks to stand in the L-shaped pattern. Similarly, the side-by-side formations observed were generally due to a straight length of bench, which influenced their positioning, as they stood against it, next to each other. These physical constraints will shift and change in the ongoing trajectory of the cooking. Such constraints in the local environment, then, will be an important consideration for us in the context of potential distributed arrangements for these activities, in particular as they are accommodated over time.

Secondly, what we see in these new and varied forms of spatial configurations are a broader set of collaborative goals and activities and wider set of roles and ways in which people choose to participate. For example, people may work together on a task, they might perform separate activities in parallel, one might be showing others how to cook through visual demonstration or by guiding hands on-learning from behind. For some, the cooking is a focal point of the interaction, for others, it is an incidental part of being together – a context in which social interaction and conversation are managed even when attention is focused elsewhere. What we see in the analysis is how some of the particular spatial arrangements we observed were, in part, contingent on these activity characteristics wide, e.g. V-shaped formation (when two cooks are working at a long top bench) or spooning when teaching through hands-on learning.

The implications for design from these insights are subtle but important. Our study exemplified that cooking together is a somewhat complex activity and that installing a simple videoconference system in the kitchen will not cater for the rich social interactions taking place around food preparation that makes it a meaningful and positive experience. The identification of new formations when cooking together has inspired us to rethink the kinds of positioning, images and audio that need to be supported if we are to create a blended cooking experience for people distributed geographically.

In terms of camera and screen positions in kitchens, we are inspired to move away from the traditional fixed wide-angle and stovetop views towards more dynamic and intimate views supporting the shared cooking experiences that we observed. In particular, the Spooning formation draws us to consider technology configurations capable of reproducing close-up over-the-shoulder views to provide the intimate and informative aspect seen by the person behind. From the reverse L and Z-formations we noticed that in kitchens preparation and social interaction are interwoven, not discrete activities, and that although the cooking activity and kitchen architecture may draw cooks away from typical conversational formations, cooks continue talking to each other, even when facing opposing directions.

These observations indicate that any technology supporting the flow of activity and conversation could not do so from a single vantage point if it is to reproduce the intimacy of the person-to-person interaction. Audio sampling would need to take into account current body and head positioning of cooks, and a constant digital feed of images of the food, the cooks or both may not even be necessary. At the same time, we can see that moving about, working on different cooking tasks and swiveling your head to engage with the other person works well in particular patterns within the kitchen space. The implication being that given a particular spatial kitchen design it should be possible to predict locations for screens to provide an effective blended cooking experience. These are design implications that we will explore in the future, and that we encourage others to consider too.

5.2 Digital Ethnography for Design

Methodologically we have explored a "digital ethnography" approach to gaining insight into activities in everyday life in people's private homes through user-created videos posted on YouTube. While this is not the first study in HCI research to do so, we still think it is worth reflecting a bit on this approach to empirical studies in domestic settings for the purpose of informing interaction design.

Two things one could criticize the YouTube approach to ethnography for are that 1) it might not provide a representative picture of the behavior of the general population, and 2) the presence of the camera may cause people to behave differently than they normally would. These are valid concerns and certainly warrant further thinking. They are, however, not new concerns pertaining only to the use of YouTube videos, but known issues in empirical research for domestic computing. Our perspective on the subject is that in order to make reasonable judgment about methodological appropriateness one needs to consider the *purpose* of the study. From this perspective, many empirical user studies in HCI research differ considerably from empirical studies of people in, for example, sociology and anthropology, where ethnography has its origins. In contrast, HCI researchers might not be aiming at theory about people and society as such, but instead "just" be interested in informing and inspiring interesting and novel interaction design. In this case, traditional ethnography might be a hindrance rather than an enabler of innovation. To deliberately focus on "extreme" users rather than mainstream (the IDEO approach) can sometimes be more useful for design. In our experience from this study, we found that using YouTube as our data source enabled us to gain understanding and inspiration valuable for future design. We were able to observe a broad cohort of people doing interesting social interaction in their kitchens beyond what we possibly could have achieved in a similar time frame by going into the field. Furthermore we were able to do purposeful sampling [30] and focus on particular observations and emergent themes as the study evolved.

In terms of the camera influencing people's behavior it was our clear impression that this was *not* a threat to validity. On the contrary, the obvious presence of the camera sometimes actually added to the insight gained, as people would be *using it actively* in communicating with the remote viewers – something in itself of great importance when interested in supporting distributed social cooking. One could say that the presence of cameras in people's kitchens was so overt that it was almost covert.

5.3 Designing Technologies for Cooking Together

The implications for design from these insights are subtle but important. Our study exemplified that cooking together is a somewhat complex activity and that installing a simple videoconference system in the kitchen will not cater for the rich social interactions taking place around food preparation that makes it a meaningful and positive experience. The identification of new formations when cooking together has inspired us to rethink the kinds of positioning, images and audio that need to be supported if we are to create a blended cooking experience for people distributed geographically. In terms of camera and screen positions in kitchens, we are inspired to move away from the traditional fixed wide-angle and stovetop views towards more dynamic and intimate views. In particular, the Spooning formation draws us to consider technology configurations capable of reproducing close-up over-the-shoulder views to provide the aspect seen by the person behind. From the reverse L and Z-formations we noticed that cooking and conversations are interwoven, not discrete activities, and that although the cooking activity and kitchen architecture may draw cooks away from typical conversational formations, they continue talking to each other, even when facing opposing directions. These observations indicate that any technology supporting the flow of activity and conversation could not do so from a single vantage point if it is to reproduce the intimacy of the person-to-person interaction. Audio sampling would need to take into account current body and head positioning of cooks, and a constant digital feed of images of the food, the cooks or both may not even be necessary. At the same time, we can see that moving about, working on different cooking tasks and swiveling your head to engage with the other person works well in particular patterns within the kitchen space. The implication being that given a particular spatial kitchen design it should be possible to predict locations for screens to provide an effective blended cooking experience. These are design implications that we will explore in the future, and that we encourage others to consider too.

Acknowledgements. We would like to thank Adam Kendon for valuable comments on our research.

References

1. Bell, G., Kaye, J.: Designing technology for domestic spaces: A Kitchen Manifesto. Gastronomica, 46–62 (2002)
2. Blythe, M., Cairns, P.: Critical Methods and User Generated Content: the iPhone on YouTube. In: Proc. CHI 2009, pp. 1467–1476. ACM (2009)
3. Blythe, M., Cairns, P.: Tenori-On Stage: YouTube As Performance Space. In: Proc. NordiCHI 2010, pp. 72–81. ACM (2010)
4. Bonanni, L., Lee, C.-H., Selker, T.: Cooking with the Elements: Intuitive Immersive Interfaces for Augmented Reality Environments. In: Costabile, M.F., Paternó, F. (eds.) INTERACT 2005. LNCS, vol. 3585, pp. 1022–1025. Springer, Heidelberg (2005)
5. Crabtree, A., Rodden, T.: Domestic Routines and Design for the Home. Computer Supported Cooperative Work 13(2), 191–220 (2004)
6. Gaver, B., Dunne, T., Pacenti, E.: Design: Cultural probes. Interactions 6(1), 21–29 (1999)
7. Greenberg, S., Marquardt, N., Ballendat, T., Diaz-Marino, R., Wang, M.: Proxemic interactions: the new ubicomp? Interactions 18(1), 42–50 (2011)
8. Grimes, A., Harper, R.: Celebratory Technology: New Directions for Food Research in HCI. In: Proc. CHI 2008, pp. 467–476. ACM (2008)
9. Hall, E.T.: The Hidden Dimension. Anchor Books (1966)
10. Hamada, R., Okabe, J., Ide, I., Satoh, S., Sakai, S., Tanaka, H.: Cooking Navi: Assistant for Daily Cooking. In: Proc. Multimedia 2005, pp. 371–374. ACM (2005)
11. Hamer, B.: Kitchen Stories (Salmer fra kjøkkenet) [Motion picture]. Norway (2003)

12. Harley, S., Fitzpatrick, G.: Creating a conversational context through video blogging. Computers in Human Behaviour 25(3), 679–689 (2009)
13. Heaton, L.: Preserving Communication Context: Virtual Workspace and Interpersonal Space in Japanese CSCW. In: Proc. Cultural Attitudes Towards Communication and Technology 1998, pp. 207–230. University of Sydney, Australia (1998)
14. Hung, H., Kröse, B.: Detecting F-formations as Dominant Sets. In: Proc. ICMI 2011, pp. 231–238 (2011)
15. Huettenrauch, H., Eklundh, K., Green, A., Topp, A.: Investigating Spatial Relationships in Human-Robot Interaction. In: Proc. IROS 2006, pp. 5052–5059 (2006)
16. Hutchinson, H., Mackay, W., Westerlund, B., et al.: Technology probes: inspiring design for and with families. In: Proc. CHI 2003, pp. 17–24. ACM (2003)
17. Ju, W., Hurwitz, R., Judd, T., Lee, B.: CounterActive: An Interactive Cookbook for the Kitchen Counter. Ext. Abstracts CHI 2001, pp. 269–270. ACM (2001)
18. Kendon, A.: Conducting Interaction: Patterns of Behavior in Focused Encounters. Cambridge University Press, Cambridge (1990)
19. Kendon, A.: Spacing and orientation in co-present interaction. In: Esposito, A., Campbell, N., Vogel, C., Hussain, A., Nijholt, A. (eds.) COST 2102 Int. Training School 2009 LNCS, vol. 5967, pp. 1–15. Springer, Heidelberg (2010)
20. Lagomarsino, A.: Kitchen album: concept based on progressive user research. In: Proc. DPPI 2007, pp. 423–428. ACM (2007)
21. Marshall, P., Rogers, Y., Pantidi, N.: Using F-formations to analyse spatial patterns of interaction in physical environments. In: Proc. CSCW 2011, pp. 445–454. ACM (2011)
22. Masten, D.L., Plowman, T.M.P.: Digital Ethnography: The next wave in understanding consumer experience. Design Management Journal 14(2), 75–81 (2003)
23. Mennicken, S., Karrer, T., Russell, P., Borchers, J.: First-Person Cooking: A Dual-Perspective Interactive Kitchen Counter. Ext. Ab. CHI 2010, pp. 3403–3408. ACM (2010)
24. Mentis, H., O'Hara, K., Sellen, A., Trivedi, R.: Interaction proxemics and image use in neurosurgery. In: Proc. CHI 2012, pp. 927–936. ACM (2012)
25. Morrison, C., Jones, M., Blackwell, A., Vuylsteke, A.: Electronic patient record use during ward rounds: a qualitative study of interaction between medical staff. Critical Care 12(6), R148 (2008)
26. Nardi, B., Schiano, D., Gumbrecht, M., Swartz, L.: Why we blog. Communications of the ACM – The Blogsphere 47(12), 41–46 (2004)
27. Palen, L., Hughes, A.: When home base is not a place: parents' use of mobile telephones. Pers. Ubiquit. Computing 11(5), 339–348 (2007)
28. Obata, A., Sasaki, K.: Office Walker: a virtual visiting system based on proxemics. In: Proc. CSCW 1998, pp. 1–10. ACM (1998)
29. O'Hara, K., Kjeldskov, J., Paay, J.: Blended interaction spaces for distributed team collaboration. ACM TOCHI 18(1), article no. 3 (2011)
30. Patton, M.: Qualitative Research & Evaluation Methods. Sage Publications, Thousand Oaks (2002)
31. Rotman, D., Preece, J.: The 'WeTube' in YouTube – creating an online community through video sharing. Int. J. Web Based Communities 6(3), 317–333 (2010)
32. Rogers, Y.: HCI Theory: Classical, Modern, and Contemporary. Morgan & Claypool (2012)
33. de Runa, J., Harpring, J., Rafiuddin, M., Zhu, M.: Not Enough Cooks in the Kitchen. In: Proc. CSCW 2010, pp. 485–486. ACM (2010)
34. Terrenghi, L., Hilliges, O., Butz, A.: Kitchen Stories: sharing recipes with the Living Cookbook. Pers. Ubiquit. Computing 11(5), 409–414 (2007)
35. Tran, Q.T., Calcaterra, G., Mynatt, E.D.: Cook's Collage: Deja Vu Display for a Home Kitchen. In: Sloane, A. (ed.) HOIT 2005. IFIP, vol. 178, pp. 15–32. Springer, Boston (2005)

Practices Surrounding Event Photos

Dhaval Vyas[1], Anton Nijholt[2], and Gerrit van der Veer[3]

[1] Industrial Software Systems, ABB Corporate Reserach, Bangalore, India
[2] Human Media Interaction, University of Twente, The Netherlands
[3] School of Computer Science, Open University, The Netherlands
dhaval_vyas@yahoo.com, a.nijholt@utwente.nl, gerrit@acm.org

Abstract. Sharing photos through mobile devices has a great potential for creating shared experiences of social events between co-located as well as remote participants. In order to design novel event sharing tools, we need to develop in-depth understanding of current practices surrounding these so called 'event photos'- photos about and taken during different social events such as weddings picnics, and music concert visits among others. We studied people's practices related to event photos through in-depth interviews, guided home visits and naturalistic observations. Our results show four major themes describing practices surrounding event photos: 1) representing events, 2) significant moments, 3) situated activities through photos, and 4) collectivism and roles of participants.

Keywords: Photo-sharing, HCI, Design, Ethnography.

1 Introduction

In the last decade, the notion of domestic photography has emerged to describe a set of photographic activities, including capturing, recording, and sharing, of ordinary people creating and utilizing images for non-professional purposes [23]. To Slater [24], it is a constructed act, where a specific area is located on the camera viewfinder and people give a particular pose to the camera. After the capture, photos are (collaboratively) selected, uninteresting photos are kept aside, and a selection of photos are shared, archived or even framed in people's homes. Photos created in such a way help people construct a positive image, individual or family identity and support their emotional, sentimental and nostalgic needs [23].

Sharing events through photos has become a genre within domestic photography. With the advent of smartphones and related portable mobile technologies, it has become easier to share photos and other relevant media such as music, videos and texts while on the move. Mobile phone based event sharing applications such as Vivogig, Google+ Events, Color and Facebook Events are a few examples of existing tools. In the HCI community, a number of studies have contributed to the genre of event sharing. For example, Jacucci et al. [14] studied how participants using multimedia-capable mobile phones collaboratively shared their experiences at a large-scale car racing event. Brown and Chalmers [5] studied practices and behaviors of tourists, in order to provide design implications for novel event sharing tools (related to visits and

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 55–72, 2013.

travels). George Square [6], MySplitTime [12], Automics [11], and Columbus [22] are a few examples of tools that support event sharing experiences via photos.

In this paper we focus on 'event photos', as an integral part of domestic photography. We define events as any spontaneous or planned activities that have some level of social meaningfulness, for example, weddings, birthday parties, vacation trips or picnics, amongst others. Photos taken during such events or about such events are termed as event photos. The context and social meaningfulness are central to event photos. We use the notion of 'events' as a lens to understand people's practices surrounding domestic photography and inform the design of novel technologies that can support event sharing. As such, we are not interested in differentiating event photos from non-event photos.

To understand the importance of event photos and how they are dealt with, we carried out a field study. We applied three different methods in our study: in-depth interviews with 17 participants, guided home visits to 9 families, and naturalistic observations at 4 event locations. In this paper, we will provide a detailed analysis of our study and describe our results in four themes describing practices surrounding event photos: 1) representing events, 2) significant moments, 3) situated activities through photos, and 4) collectivism and roles of participants. Using results of our study, we also provide implications for designing novel event sharing tools.

1.1 Motivation: Why Event Photos?

As we started our research on event sharing via mobile devices and relevant technological explorations, we realized that there was a lack of literature on how photos were taken during events and what features of such events were captured more than others. There was certainly enough material on photo archiving, sharing and curating and also on sharing via mobile devices (see [23] for an up-to-date review). However, to what extent this literature could inform the design of innovative event sharing tools for mobile devices was not clear to us. For example, what are the motivations for taking photos when people go to events – both public and private? What strategies do people use when capturing event photos? How are event photos shared, and what are the motivations for sharing such event photos? Similarly, how significant are these event photos in people's everyday lives and how are they curated and archived for long term use?

One might argue that the majority of photos that people capture and share may already have relevance to some kind of events. Hence, what value may such an explicit focus on event photos bring to the HCI community? We believe that there are several reasons why event photos need an explicit focus:

Different events have varying social order and social dynamics. Some of these orders and dynamics are formal (e.g. wedding) and others are casual and ad-hoc (friends meeting at a pub or a picnic place). These events affect the nature of photos and the role of a photographer and people who are taking part in the event. For example, one might be less motivated to bring a large-sized DSLR camera to a pub gathering of friends. One might just walk-in without any extra effort knowing that a camera phone would be sufficient for such an event. This not only shows the significance of the event but also people's attitude and expectations from such events.

Events also have their moments of significance that can be worth capturing and without which the event may lose its integrity or value. For example, candle blowing at a young child's birthday party is an important moment that should be captured. Similarly, at a wedding, the exchange of rings or the first dance could be seen as significant moments. Depending on different cultures, these significant moments may change but the tendency to capture such moments is pervasive.

While taking photos at an event, one is both experiencing and capturing a moment at the same time. Elements of "lived" experiences may not be easily conveyed when these photos are shared via email after the event is finished. The situatedness of events shape the way photos are captured, shared and interacted with. Similarly, captured photos may also affect people's activities and behavior. Recent smartphones do allow sharing of photos immediately after they are captured via, for example, social networking services (SNS). But is that sufficient for sharing an event? Also the photos that are captured during an event can be seen as collective contents [20]. There might be some interesting privacy issues that need to be explored.

We believe that there are many such issues that can be answered by a better understanding of event photos. A field study in this area can allow us to assess current practices and methods people apply in capturing, sharing and using event photos for a longer-term.

2 Related Work

2.1 Photos in Homes

Chalfen's [7] work pointed out that the primary value of domestic photography is related to memory, identity and communication. He coined the term Kodak Culture, referring to the traditional family practice of sharing printed photos or video footage of friends and family in a collocated environment. The HCI community has since seen a rich body of work in this domain, focusing on organizing, sharing and displaying photos in homes. Using observations and interviews, Kirk et al. [16] developed the notion of photowork that describes different sets of activities that users perform on their digital photos after the capture but prior to sharing. Photowork highlights the efforts and complexity of practices such as reviewing, downloading, organizing, editing, sorting and filing. Focusing on the sharing aspects, Frohlich et al. [13] developed requirements for photoware – an application that supports storage, sending and sharing of digital photos. Crabtree at al. [9] studied conversation patterns around collocated photo sharing with a goal to support distributed collaboration. Recently, some ethnographic studies have also looked at how families curate and display photos around their homes. A set of studies from Microsoft Research [10, 25, 26] has illustrated how different physical setups (e.g. walls, mantelpiece and shelves) in the homes are utilized by family members to display important photos and how such displays are tightly interwoven into the very notion of family and home. Drawing from four different studies, Van House [27] showed different ways people share photos in their home, using prints, photo albums, slideshows, via laptop screen and so on. In her later work [28], she describes personal photography as overlapping technologies of memory, relationships, self-representation, and self-expression. A journal special issue on ‘collocated social practices surrounding photos’ [17] elicited three broad themes: reflection and remembrance; performativity and expression; and connection and communication.

2.2 Photos in the Online World

Echoing the storytelling aspect of photos in the offline world, Balabanovic et al. [3] explored two categories of methods people used in telling stories from digital photos: photo-driven and story-driven. Using semi-structured interviews, Miller and Edwards [19] studied digital photo sharing practices of 10 participants on Flickr. They explored two categories of users: people who were still following the Kodak Culture and ‘Snaprs’. Snaprs are the ones who shared their photos even outside of their social network with fewer concerns for privacy. Their immediate focus was on taking photos then sharing them to relevant people. Ahern et al. [1] identified four factors that could affect people’s privacy while sharing digital photos: security, identity, social disclosure, convenience. Bentley et al. [4] compared personally captured photos to commercially purchased music and found several similarities. From this comparative study, they found out that 1) users search with fuzzy concepts and settle for an “okay” option, and 2) users change their mind during the search process and end up with something completely different.

2.3 Photos in Mobile Devices

One of the earliest studies on photo sharing through mobile devices was published by Mäkelä et al. [18]. These researchers identified that people shifted from telling stories about the pictures, to telling stories with the pictures. Kindelberg et al.’s [15] study of camera phone users led to taxonomy of six affective and functional reasons for image capture on a camera phone: individual personal reflection, individual personal task, social mutual experience, social absent friend or family, social mutual task, and social remote task. Ames et al. [2] have provided detailed requirements for mobile photoware. Olsson et al. [21] studied users’ needs for sharing the digital representations of their life memories. They identified three main motivations: personal growth and identity (no sharing), strengthening social ties (sharing with family and friends) and expressing/getting attention (sharing with anyone). In a field study, Jacucci et al. [14] explore how people actively construct experiences using mobile devices capable of sharing multimedia content. In particular, the authors suggest that continuity, reflexivity with regard to the self and the group, maintaining and re-creating group identity, protagonism and active spectatorship were important social aspects of the experience. Another such event sharing study that comes close to our own research interests was done by Esbjörnsson et al. [12]. From an ethnographic study at car racing venues in the UK and Sweden, the authors describe of three interesting findings that can be useful for supporting event sharing at car racing venues: viewing paradox of spectating, active spectating and role of sociability. Brown and Chalmers [5] also use the notion of events, where they study behavior and practices of tourists in order to derive implications for tourism related tools. Automics [11] is a mobile application that allows users to collaboratively generate photo collages using photos from different sources in a theme park setting. Columbus [22] provides a location-centric event sharing experience, in which, by walking around different locations users can unlock photos that are digitally associated to different locations. Adding voice and location modality to photos, the George Square [6] system allows users to access historical data from previous visits to be able to collaboratively explore physical places.

3 Field Study

We used three methods to study event photos: 1) in-depth interviews, 2) guided home visits, and 3) naturalistic observations during real events. For the in-depth interviews, we created an online survey to collect information about our potential participants' interests and experience in photography, their current practices and their availability. Our survey was completed by 40 participants out of whom 17 were invited for in-depth interview sessions. We selected these participants based on their diversity of experiences and photo taking practices. Details of our participants are in table 1a. In the follow-up interviews, figure 1 (left), participants were asked to bring at least 10 photos from three of their recent social events. During the interview, we first asked questions about their capturing, sharing and archiving practices related to photos. In the second part, we asked our participants to describe the three events, using photos that they had brought. Some example questions are provided in the following:

- What tools have you used for capturing, storing and sharing these images?
- What was your aim and motivation behind taking this picture?
- What does this picture signify in your life?
- Were these pictures ever discussed with others? How?
- How do you classify these photos?
- Do you use any of these pictures as memento? Why?

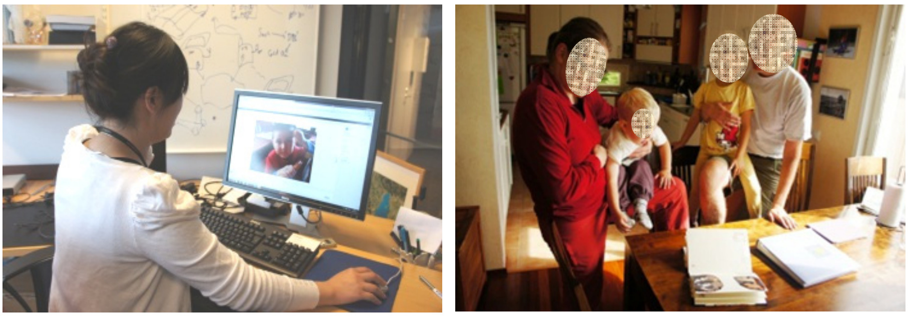


Fig. 1. A participant describing her event photos in the interview session (left) and a family showing their photo albums in the home visits (right)

Table 1. Details of our participants in interview (a) and home visit sessions (b). § We had involved all five participants in this session including the grandparents of the family (#4) who were staying with them at the time of our visit.

In-depth Interviews	
#	Gender/Age
1	Male/34
2	Male/30
3	Female/30
4	Female/24
5	Female/26
6	Female/32
7	Female/44
8	Male/48
9	Male/30
10	Male/44
11	Male/27
12	Female/34
13	Female/31
14	Male/28
15	Male/21
16	Female/47
17	Male/40

(a)

Guided Home Visits	
#	No. of Members
1	5: 1 – Father; 1 – Mother; 3 – Daughters
2	3: 1 – Father; 1 – Mother; 1 – Daughter
3	4: 1 – Father; 1 – Mother; 2 – Daughters
4	5§: 1 – Father; 1 – Mother; 1 – Daughter; 2 – Grand parents
5	3: 1 – Father; 1 – Mother; 1 – Son
6	2: 1 – Mother; 1 – Daughter
7	4: 1 – Father; 1 – Mother; 1 – Daughter; 1 – Son
8	4: 1 – Father; 1 – Mother; 1 – Daughter; 1 – Son
9	2: 1 – Husband; 1 – Wife

(b)

With the permission of our participants we audio recorded these interviews and used their photos for our analysis. These participants were given a gift at the end of the interview session as a token of appreciation.

In the guided home visits, we recruited 9 families, from the Helsinki region of Finland, using a similar kind of survey as we used for our interview sessions. The overall process of these home visits was very similar to our interview sessions. But here we also wanted to explore their curation and display of printed photos. We also used a similar approach in discussing the event photos – describe 3 events using at least 10 photos. In some cases, families used their computers to describe these events and in others they used their physical albums and other printed photos. Figure 1 (right) shows a family describing an event using their physical albums. The details of our home visit participants are provided in table 1. The complete home visit was audio recorded and a large number of photos were captured depicting interesting photo archiving and displaying practices. These families were also given a gift at the end of the interview session as a token of appreciation.

Thirdly, naturalistic observations in real events were carried out in a very informal way. We visited 4 places (a zoo, an amusement park, a cultural event and a public park) to observe how families and groups of friends take pictures. At times, we approached some of the visitors who were taking pictures and asked them about their motivations behind taking such pictures and how they would share them with others. We took notes whenever it was necessary. Although these observations were very informal and ad-hoc, our intention here was to gain naturalistic insights into events.

At the end of our field study, we had collected a large amount of field data in the form of notes, photos and audio recordings. In addition, we also used event photos that were discussed by our participants at the time of the interview and home visits. We did a qualitative data analysis and used affinity wall exercise to explore patterns and after discussing with other members of our team, we developed our results.

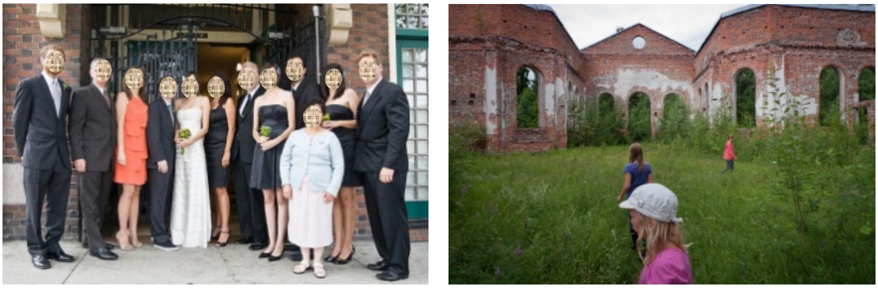
4 Findings

In the following, we provide findings from our field study focusing on four major themes that came out strongly in our analysis of the collected data. These themes are representing events through photos, capturing significant moments during events, situated activities through photos and collective effort and roles of people.

4.1 Representing Events

Based on our interview and home visit data, we found that representing and documenting actual events in the form of photos was one of the most visible practices. Participants attempted to represent relevant activities that happened during events through a set of photos. We noted that formal events such as weddings, baptizing of a child, and wedding anniversaries were always well documented via photos. The main reason for documenting such events was that these events signify a great value in their lives and they could go back to these event photos as important memories of celebrations and achievements.

From the field study, we found that our participants used some interesting strategies while they captured photos during events. One of the most basic strategies was to capture the relevant persons (either in group or individually) who were part of a social event. In particular, capturing ‘group photo’ was a widely used strategy to represent who were present at a given event. Figure 2a shows an example of this strategy, where an interview participant (#2) who attended his brother’s wedding, took a photo of the two main families for remembering the event. In contrast to the above example, an interview participant (#1) who went on a picnic trip with his friends from a Facebook based social club commented: *“I have been to many trips with friends whom I meet only casually. I would always suggest taking a group photo on the first day or at the time when people meet for the first time. You never know, on the last day things may have gone awkward and you may never get a group photo.”* The above two examples show how group photos serve different purposes for formal and informal events.



(a) (b)

Fig. 2. (a) A group photo, capturing important people at a wedding and (b) Photo of an old church for documenting a trip

Capturing photos of different activities during events was also a strategy used by our participants to represent their experiences in events. During formal events, such as weddings, the intention would be to capture the sequence of all important activities. On the other hand, lesser significant events such as going on a vacation trip, visiting an amusement park or going to a concert were documented with less focused intentions. For example, figure 2b shows a photo from a participant from our home visit (#1) who attempted to document a broken church that his family saw during their trip to Russia. Even though the photo did not have any strong significance to the trip, the participant took this photo mainly for documentation purpose.

The nature of events also played an important role in our participants' motivations for taking photos. In some cases, participants who had plans to visit a new place (during vacation or work related visits) took their DSLR cameras with the expectation that they might see something interesting in a new place. On the other hand, participants' attitudes differed when they visited known places. Here is a comment by one of our participants (#9) from the interview sessions; *"I wouldn't bother to bring a camera with me if I am going to visit my friends in Turku* (~165 km from Helsinki). *Because I know that place very well and I know that nothing interesting would happen when I am there."*

Another interesting photo capturing strategy was to take photos depicting a starting point and an end point – hence enabling storytelling (Van House, 2009). During the field study, all the participants started describing their events from a photo that was taken at the beginning of an event and finished their conversation at a photo that described the end of their event. In the interview, we often heard our participants saying *"this is how we started the trip"* and similarly, *"this was the last day of our vacation"*, for example. This finding shows that all events carried a story and a narrative that was always conveyed through their photos. This practice was observed in almost all participants and it was a useful way of not only documenting but also remembering an event.

4.2 Significant Moments

The second theme that came out strongly in our study was how participants attempted to make sure that the significant moments and aspects of events are always captured. Moments such as blowing candle on a birthday cake, first dance at a wedding or giving toast at an anniversary party are obviously important and can be termed as central to defining such events. Similarly, during non-formal events such as vacation trip or any other type of outing, there are moments that define or characterize the whole trip. In this section, we will describe our participants' motivations and practices related to such important moments in events.

Participants captured photos of novel, rare and significant aspects of events to shape their experiences related to events. As we mentioned earlier, some events (e.g. a wedding) have a defined social order and there are observable moments of significance such as exchanging rings or a wedding speech. These significant moments were very carefully captured. Figure 3a shows an example of a significant moment at a 50th wedding anniversary celebration of our participant's (#11) parents, where he took a photo of his elder brother giving a gift and making a toast to their parents.

(a) (b)

Fig. 3. Examples of significant (a) and iconic (b) moments during different events

Figure 3b shows an example where our interview participant (#2) had gone on a trip to Africa with his friends. At the end of the trip (which was described as very "hectic" and "difficult", in his words) all the three friends had lost a considerable amount of weight and were very relieved to go home. Figure 3b shows an iconic photo of the three friends traveling in a boat to catch a flight back to the US. Our participant gave the following comments on the photo: *"this was an iconic moment of our trip. It summarizes our whole trip. It shows how tired we were and how much weight we had lost in that short trip… I took this photo when my friends were resting. And they were quite surprised when I sent them this image after reaching the US."*

The significance of events and people in them make a photo important enough for sharing and archiving for a long term use. Figure 4a is a postcard created by our interview participant (#12). The original photo was taken during a holiday visit with three of her friends, who got together after a long time. After the holiday, a postcard from this photo was created and was sent to the other two friends to cherish memories of that event. The participant commented: *"this photo represented a nice holiday with* *my best friends, so I thought a post card would make it even more memorable."* From the home visits, we found that photos related to children were frequently displayed on living room walls, fridges, mantels and desks. Most of the families (#1 to #8) who had children kept one or more photo albums for collecting childhood memories for their children and for record keeping purposes. Figure 4b shows a page in a photo album, which shows a set of photos from a baptism ceremony of a child belonging to the family (#7). During our home visit session, the mother of this family indicated that she has made detailed annotations about this event on the album itself. She stated that, *"Without annotations things can be forgotten."*

(a) (b)

Fig. 4. A postcard created from an event photo (a), and Photo album depicting a baptism ceremony (b)

4.3 Situated Activities and Photos

We observed that certain activities and interactions during events led to capturing and sharing of photos and the other way around. The effect of this theme was observed during events as well as after events. In this section, we will provide examples from our participants who supported interaction through photos and vice versa.

In the following we provide a short account from a family who participated in our home visits (#1). The family organized a trip to Russia with other close relatives to see their ancestral birthplace. Our participant gave the following account about this event: *"My father is 80 and is not well at the moment. He asked me and my brother to go to his birthplace which used to be a part of Finland before the war with Russia, but now is in Russia. He couldn't visit that place after the Soviet Union came. My father gave us some old photos of that place and told us to go to his village and get photos of its current look. We were accompanied by our uncle who knew a bit about this place. When we arrived at the place, we did some sightseeing around the village, saw the building where he lived, went to a church nearby that was completely deprived and some of the places that my father often referred to. We could almost compare the photos he gave us with the current state of the location. We also saw our family names written in that church. While we were walking around the place, our uncle used these old photos almost like a map and guided us through different* *locations in the village. By comparing the old photos to the current situation he also told us lots of stories about the place: what they did when they were kids and what their life was at that time. My brother and his kids also took a lot of pictures around the area."* The example shows how photos from the past aided and guided the family to look around the place and imagine how the life was fifteen years ago. Importantly, as mentioned by the participant, these photos served as a "map" and itinerary of different locations that the family might visit. This example show how old photos supported tourist activities and how the parts of the trip were guided by these photos.

There were also examples where photos captured during an event led to interaction within groups. In the interview session a participant (#17) gave an account of his family trip to Lake Garda in Italy. The family had decided to go to two different places during a day, where the father would accompany their son and the mother would accompany two daughters. The following interview excerpt shows how using MMS the groups shared their photos, which led to photo based interaction. *"Since we had a lot of ground to cover and had just one more day remaining in our trip, we thought to go to different places around Lake Garda. After a few hours I got an image from my wife where one of my daughters is shown with a teddy bear. She then immediately called me and told me that Mia – our daughter has won a teddy bear in a shooting game at a local gathering. When I showed it to my son he also urged me to go to a similar place. Unfortunately, at our location there wasn't anything like that. So, I brought an ice-cream for him and sent his photo with the ice-cream to my wife's phone. This way of conversation happened a couple of times."* This excerpt shows how the situatedness of photos allowed the two groups to communicate and communicate fun at each other.

Central to this theme is the phenomenon of how "lived" experiences are shaped through photos. The example of the family trip to Russia shows how old photos given by the ailing father served as a tool not only to navigate through the village but also worked as an experiential artefact [29] which allowed the family to experience the trip from the perspective of their father. The example where a family splits into two groups highlights how lived experiences are shared by sharing photos via mobile phones.

4.4 Roles and Collectivism

One of the interesting aspects that came out from our field study was the roles different people played within the context of domestic photography. Sometimes the role of a photographer was explicitly assigned and sometimes it was assumed. For example, a wife who participated in our home visits (#9), mentioned that once she was asked to be a maid-of-honor at her close friend's wedding, which implicitly included the duty of taking photos during her friend's bachelorette party. In another case, an interview participant (#3) was asked to take photos at a social event where she was also part of the organizing committee. In different cases, tech savvy members of families were simply assumed to take care of photo capturing during events.

Collectivism was another aspect that came out very strongly in our work. In several examples, we observed that photo capturing was divided between different people in the event. In the example where a family went to Russia to visit their ancestral place, several family members including children had cameras with them. The participant that we interviewed mentioned, *"At this moment, I am collecting photos and stories related to them, from all the family members and planning to make a photo-book out of this trip. If we go back to the same place again in ten years, we won't have anybody telling stories about this place, so I think it's a good idea to make an album of this trip as a family heirloom."* This account indicates two characteristics of event photos: social meaningfulness of the event itself and people's role in the event. The event itself did not have any formal nature compared to a wedding ceremony, for example. However, the place had a great emotional relevance for the family. The account hints at the roles of people in creation of these event photos. It showed how several persons were taking photos on the trip, how the uncle was guiding the group and their photo-taking activities and how the group agreed to collect photos from all sources to create an album.

Reciprocity was seen as central to the idea of collectivism. An interview participant (#6) mentioned about her trip to Norway, where all the six friends had a camera with them. During the trip each of the members took photos and at the end of the trip they decided to share all the photos. After reaching home, our participant (#6) created a Picasa account where she shared her photos from the event and invited others to share their photos. Figure 5 shows a screenshot of that participant's Picasa gallery. One can see 6 different albums from the same trip uploaded by the 6 friends who went on the trip together. In this case the authorship and ownership of the photos were shared. This was one of the few cases where such an open reciprocity was shown.

Fig. 5. A Picasa gallery screenshot

In several cases, however, participants sent and received only selected photos via emails or via online photo galleries. Some participants even raised a concern about not receiving interesting images from the people who were on the same trip. Highlighting this fact, one interview participant (#11) commented: *"If you don't collect your photos just after the event finishes, most likely you won't get the photos later."*

5 Discussion and Implications

In this paper, we have used the notion of 'event' as a lens to understand different practices related to domestic photography. Event photos emphasize the social meaningfulness and the dynamics of events. We found that the nature of events, to a large extent, shaped the motivations and practices surrounding domestic photography.

The notion of sharing in private groups emerges from our observations: when a group of friends or family members take part in a social event, they often collaborate and share their photos during events and even after the events. The roles of individuals became visible in the activities of taking event photos, for example, the 'assumed' and 'assigned' role of photographers in an event. The plurality of photo-takers shows how they cooperate and exchange photos with each other. While capturing event photos, people are motivated towards documenting an event, keeping in mind its social order of the event (e.g. wedding); they also make sure that they have every individuals' photos taken (e.g. group photo); and particularly focus on significant or iconic moments from events. As a drawback, they often end up with segmented photo collections with a lack of shared knowledge about photos captured by different people. We observed that often one person took the responsibility of making sure that photos were shared. But such reliance on explicit user actions was time consuming. As we noted from our results, it is difficult to keep track of photos taken by different people during a trip or event. The case discussed in figure 5 seems to be the best-case scenario where everyone is obliged to send their versions of photos. It may not always be easy to share photos once the event is completed. Asking for these photos again and again from individual photographers might even affect social relationship of people.

Our participants were cautious about how they shared photos to a large audience. For event photos, private channels such as email or links to secured photos (e.g. Picasa) were preferred over SNS such as Facebook and Google+. Important photos were printed in different forms (e.g. postcards, albums) to convey a sentimental value associated with such event photos.

Existing literature on domestic photography focuses on aspects such as storing and sharing [13], photo work issues [16], storytelling, identity and other personal aspects [9, 27]. Our findings provide directions for designing event sharing tools for mobile devices. In the following, we will lay down directions for design.

5.1 Event Sharing and the 'Cloud'

Smart phones with built-in cameras and other sensors have become a pervasive commodity these days and their hardware features and capabilities will only increase over time. Considering events in private or public space, a likely scenario is that more than one person will be able to take photos using their camera phones. We believe that by using cloud computing infrastructure new services can be designed that will allow users to share their experiences.

Following the notion of private groups, photo sharing services can be developed that could allow creating private groups while being at an event. Others can join in through ad-hoc group formation using NFC or Wi-Fi connections. All the users in a particular photo group can take photos and store them in a shared repository on a cloud, which is accessible to all the members. We can think about several possibilities for allowing automated photo sharing features that can replace the manual transfer and sharing of photos. For example, specific sharing features can be included that allow automatic sharing of photos only to those friends who were part of a particular event, disabling others to see those photos. Additionally, using advanced face recognition features, photos can be shared only to the people whose faces appear on the photo.

Novel services could also allow users to 'broadcast' their interests for specific images, where a user is notified whenever a new photo of him/her appears on the cloud, for example. This way, if the user's photo appears in the album of a complete stranger, the user can request that photo. On the other hand, she could be enabled to declare not to appreciate capturing, or sharing such photos in public domain. Users can subscribe to specific entities of events (e.g. locations, people) and get notifications of new photos when they are captured. The features discussed so far are only a few possibilities. Depending on the nature of the event users should be able to choose or define such features taking their privacy concerns into consideration.

Capturing and analyzing meta-data becomes very important in such cases. Meta-data associated with photos, such as date and time, GPS coordinates, names of recognized people or objects in the photos, orientation and activities of photo-takers, and nearby users and services can be useful for analyzing certain activities and behaviors of users. Especially, the usage meta-data (e.g. how often one visits a photo, what kind of photos people look at, which circles or groups these people belong to) can be of great value, as users' consumption of photos can produce new content that can be very valuable for enhancing social interactions via such event sharing tools. These systems require a careful design approach by taking into account users' privacy concerns. Participatory design approach used in [30] could be a useful way to take into account these concerns at an early stage of design.

Once group members share their photos in a cloud repository, the service can constantly analyze such new content for making better recommendations to its users. For example, based on the existing content about an event, the system might be able to suggest which photos are redundant or which type of photos are still needed for a complete capture of the event. Based on the photos that are already taken by other members, the system might inform and recommend users about different activities that are going on simultaneously. In this case, photos can be the raw source of information and through its detailed analysis, important notifications may be generated.

5.2 Ideas for Event Sharing Services

We envision that photos and photo sharing activities can support navigation, tourism, gaming and other comparable domains. We believe that capturing useful meta-data from the smartphone and photos can lead to innovative applications and services being created.

Domestic photography has always been a social process. New technologies can make domestic photography participatory and cooperative by involving people in the event as well as outside of the event in peer-supported navigation. The participant in our home visits session (#1) gave an interesting instance of how he and his family members went to visit a village in Russia where the participant's father was born.

This instance illustrated how peers (uncle) and certain objects (old photos) could support navigation. Elaborating on this observation we envision how the use of maps and location-based services can inform users where to go for capturing interesting photos in a tourist area. Google search already provides location-based image search, however, the aspect of real-time assistance may be worth exploring further.

There are existing augmented reality applications that allow cameras to detect and recognize buildings and other relevant objects. In addition to image recognition, GPS coordinates and activities of the photographer can be detected and utilized to further refine accuracy of locations. In such a way, meta-data of photos and photo collections can be utilized as ‘check-in’ mechanisms. For tourism purposes, an idea of shared-itineraries (figure 6) can be explored: a service that keeps track of photos taken by its subscribers and stores them with reference to location, time and other metadata and automatically generates itineraries based on these data. This way, the subscribers leave cues about their visits whenever they take photos. Over time, several categories of itineraries will develop. New subscribers can make use of such itineraries by looking at the navigation path, photos and other relevant information. Companies such as Expedia or Lonely Planet could enhance their customers’ tourism experience using such services.

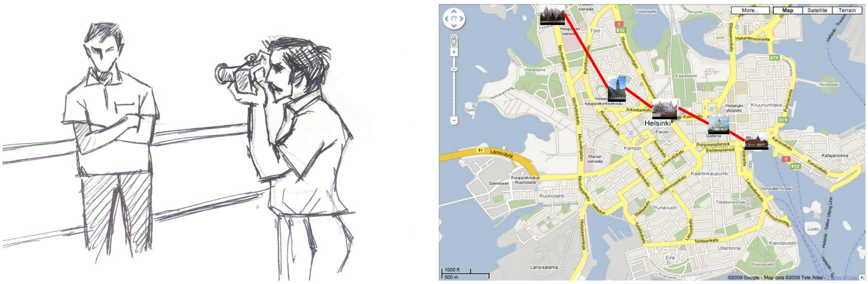


Fig. 6. Shared Itinerary concept. As subscribers take photos on their visit, itineraries are generated automatically on maps.

Photos from multiple sources could be used for providing novel location-based services: Concert organizers or amusement park owners could implement services that make use of photos from visitors who subscribe to such services. Amusement parks for children may consider search and rescue services. A cloud-based infrastructure may be utilized to collect photos of a missing child from photo streams shared by different visitors as well as from security cameras of the park. This way the system can learn a child’s roaming behavior and his/her whereabouts. An example concept could be where a dynamic 3D panorama image is generated depicting the child’s visit in different areas of the park sorted by time-stamps. Additionally, this system could be enhanced by services such as Google Maps or Twitter.

5.2.1 Designing for Significant Moments

In our findings, we observed that the appreciation of photos by our participants was related to the ‘significance’ of photos. This seems true for any type of photos.

However, it was rather difficult to generalize what made a photo significant, because the significance or value of a photo was strongly context dependent. The photo shown in figure 3a was significant because of the nature of the event and people who were in that photo, whereas the photo in figure 3b represented a shared-experience and an iconic moment that made that photo significant. It would be of great value if we could design technologies or tools that could determine significance of a photo, utilizing its meta-data. One way to extract significant photos from multiple sources in an event is by analyzing similarities in photos from multiple cameras.

Fig. 7. Plurality of photo-takers; and a graph of number of photos taken by people in an event over time, highlighting the most similar photos

Figure 7 shows a concept for exploiting the plurality of photo capture. Here, users can subscribe to a social event (e.g. rock concert, social gathering) where several people are taking photos. Users may be physically present at the event or using this service from a remote location. The concept allows users to get the most significant photos from that event in real-time. The system can collect photos taken during the same time-frame and location and make a comparison to find similarities in photos, using face recognition and/or object recognition. When such photos match with each other, then the most representative photo is identified and shared to all the subscribers of that event, as a significant moment of the ongoing event. The system can constantly look for such photos with user-generated significance (i.e. plurality of photo takers) and send these photos to the subscribers. The example graph in figure 7 illustrates how photos from multiple sources can be extracted by matching their similarities. In fact, Xerox has already developed a concept where a photo album is created in real-time from multiple sources by taking into account the plurality and similarities of photos [8]. Such an invention allows users to get the most important photos from that event in real-time.

6 Conclusions

In this paper, we provided an account of practices that surround event photos. We aimed at looking explicitly at events in order to elicit ideas for developing tools that can support event sharing via mobile devices. We believe that even though the idea of event photos is not new, our specific focus on events drew some interesting insights into the specific characteristics of event photos and how these photos are dealt with by users. Based on a field study, we developed our analysis of event photos and photo sharing paradigms focusing on four major themes describing practices surrounding event photos: 1) representing events, 2) significant moments, 3) situated activities through photos, and 4) collectivism and roles of participants. Keeping in mind the growing pervasiveness of sensor-based smart camera phones, our findings are aimed at exploring concepts and design ideas for event sharing applications.

Acknowledgements. This study was hosted by Nokia Research Center, Helsinki. We thank Wagesh Kulkarni, all of our participants and Guido Grassel for supporting this work.

References

1. Ahern, S., Eckles, D., Good, N.S., King, S., Naaman, M., Nair, R.: Over-exposed?: privacy patterns and considerations in online and mobile photo sharing. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 357–366. ACM, New York (2007)
2. Ames, M., Eckles, D., Naaman, M., Spasojevic, M., Van House, N.: Requirements for mobile photoware. Personal and Ubiquitous Computing 14(2), 95–109 (2009)
3. Balabanoviç, M., Chu, L.L., Wolff, G.J.: Storytelling with digital photographs. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 564–571. ACM, New York (2000)
4. Bentley, F., Metcalf, C., Harboe, G.: Personal vs. commercial content: the similarities between consumer use of photos and music. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Montreal, pp. 667–676. ACM, New York (2006)
5. Brown, B., Chalmers, M.: Tourism and mobile technology. In: Proceedings of the Eighth Conference on European Conference on Computer Supported Cooperative Work, pp. 335–354. Springer (2003)
6. Brown, B., Chalmers, M., Bell, M., Hall, M., MacColl, I., Rudman, P.: Sharing the square: collaborative leisure in the city streets. In: Proc. of the 9th European Conference on Computer Supported Cooperative Work, pp. 427–447. Springer, Netherlands (2005)
7. Chalfen, R.: Snapshot Versions of Life. Bowling Green State University Press, Bowling Green (1987)
8. Cifarelli, C.: Content based image selection for automatic photo album generation. Xerox Corporation US patent: 2011/0129159 A1, http://www.google.co.uk/patents/US20110129159 (last accessed on February 04, 2013)
9. Crabtree, A., Rodden, T., Mariani, J.: Collaborating around collections: informing the continued development of photoware. In: Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work, pp. 396–405. ACM, New York (2004)
10. Durrant, A., Taylor, A.S., Frohlich, D., Sellen, A., Uzzell, D.: Photo Displays and Intergenerational Relationships in the Family Home. In: Proceedings of the 23rd BCS Conference on Human Computer Interaction, pp. 10–19. BCS Press (2009)
11. Durrant, A., Rowland, D., Kirk, D.S., Benford, S., Fischer, J.E., McAuley, D.: Automics: souvenir generating photoware for theme parks. In: Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems, pp. 1767–1776. ACM (2011)
12. Esbjörnsson, M., Brown, B., Juhlin, O., Normark, D., Östergren, M., Laurier, E.: Watching the cars go round and round: designing for active spectating. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2006, pp. 1221–1224. ACM, New York (2006)

13. Frohlich, D., Kuchinsky, A., Pering, C., Don, A., Ariss, S.: Requirements for photoware. In: Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work, CSCW 2002, pp. 166–175. ACM, New York (2002)
14. Jacucci, G., Oulasvirta, A., Salovaara, A.: Active construction of experience through mobile media: a field study with implications for recording and sharing. Pers. and Ubi. Compu. 11(4), 215–234 (2007)
15. Kindberg, T., Spasojevic, M., Fleck, R., Sellen, A.: The Ubiquitous Camera: an In-Depth Study of Camera Phone Use. IEEE Pervasive Computing 4(2), 42–50 (2005)
16. Kirk, D., Sellen, A., Rother, C., Wood, K.: Understanding photowork. In: Proc. of the SIGCHI Conference on Human Factors in Computing Systems, pp. 761–770. ACM, NY (2006)
17. Lindley, S.E., Durrant, A., Kirk, D., Taylor, A.S.: Editorial: Collocated social practices surrounding photos. Int. Journal of Human-Computer Studies 67(12), 995–1004 (2009)
18. Mäkelä, A., Giller, V., Tscheligi, M., Sefelin, R.: Joking, storytelling, artsharing, expressing affection: a field trial of how children and their social network communicate with digital images in leisure time. In: The SIGCHI Conference on Human Factors in Computing Systems, CHI 2000, The Hague, pp. 548–555. ACM, New York (2000)
19. Miller, A.D., Edwards, W.K.: Give and take: a study of consumer photo-sharing culture and practice. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 347–356. ACM, New York (2007)
20. Olsson, T., Toivola, H., Väänänen-Vainio-Mattila, K.: Exploring characteristics of collective content: a field study with four user communities. In: CHI 2008 Extended Abstracts on Human Factors in Computing Systems, pp. 2967–2972. ACM, New York (2008)
21. Olsson, T., Soronen, H., Väänänen-Vainio-Mattila, K.: User needs and design guidelines for mobile services for sharing digital life memories. In: Proceedings of the 10th International Conference on Human Computer Interaction with Mobile Devices and Services, MobileHCI 2008, pp. 273–282. ACM, New York (2008)
22. Rost, M., Cramer, H., Holmquist, L.: Mobile Exploration of Geotagged Photographs. Personal and Ubiquitous Computing, 1–12 (2012)
23. Sarvas, R., Frohlich, D.M.: From snapshot to social media: the changing picture of domestic photography. Springer, London (2011)
24. Slater, D.: Domestic photography and digital culture. In: Lister, M. (ed.) The Photographic Image in Digital Culture, pp. 129–146. Routledge, London (1995)
25. Swan, L., Taylor, A.S.: Photo displays in the home. In: Proceedings of the 7th ACM Conference on Designing Interactive Systems, pp. 261–270. ACM, New York (2008)
26. Taylor, A.S., Swan, L., Durrant, A.: Designing family photo displays. In: Proceedings of the Tenth Conference on European Conference on Computer Supported Cooperative Work, pp. 79–98. Springer, Netherlands (2007)
27. Van House, N.A.: Collocated photo sharing, story-telling, and the performance of self. Int. J. Hum.-Comput. Stud. 67(12), 1073–1086 (2009)
28. Van House, N.A.: Personal photography, digital technologies and the uses of the visual. Visual Studies 26(2), 125–134 (2011)
29. Vyas, D., Heylen, D., Nijholt, A., Van Der Veer, G.: Experiential role of artefacts in cooperative design. In: Proceedings of the Fourth International Conference on Communities and Technologies, pp. 105–114. ACM (2009)
30. Vyas, D., Cui, Y., Ojala, J., Grassel, G.: *Producing while Consuming*: Social Interaction around Photos shared within Private Group. In: Nijholt, A., Romão, T., Reidsma, D. (eds.) ACE 2012. LNCS, vol. 7624, pp. 133–150. Springer, Heidelberg (2012)

Sharing Experiences over Video: Watching Video Programs together at a Distance

Anna Macaranas[1,2], Gina Venolia[1], Kori Inkpen[1], and John C. Tang[1]

[1] Microsoft Research, One Microsoft Way, Redmond, WA 98008 USA
[2] School of Interactive Arts and Technology, Simon Fraser University, Surrey, BC V3T 0A3 Canada
amacaran@sfu.ca,
{ginav,kori,johntang}@microsoft.com

Abstract. While video communication is becoming quite popular among remote friends and family, recent usage practices have been extending beyond just talking heads to remotely sharing an experience by doing an activity together. However, current video chat tools are aimed at sharing talking heads and need to be reconsidered to support remotely sharing activities. We explore a specific remote shared activity – watching video programs – through a three-phase study. We surveyed people's interest in watching video together, studied how people currently watch together in their homes, and compared different conditions for watching together in the lab. Our work explored people's current and desired practices, interactions, and technical implementations. We present our findings in themes that provide insights for designing systems that better support using video-mediated communication to share watching videos together over distance. We found that remotely watching video programs together while connected by video-mediated communication is engaging, fun, and fosters social bonds between the participants, and that these results are stronger with increased fidelity of the communication media.

Keywords: Shared experiences, proxy, telepresence, teleconferencing, video conferencing, video-mediated communication, home.

1 Introduction

Recent trends in technology have made video-mediated communication (VMC) widely available for family and friends to maintain long distance relationships. In an exploratory study on how people use VMC to stay connected, Brubaker *et al.* [6] found that distributed family and friends used video in tandem with activities such as cooking, watching TV, and even giving birth. In a similar study, O'Hara *et al.* [25] found that mobile VMC was being used for casual conversation, show-and-tell, and everyday routines (i.e., entertaining one's child during bath time). We are intrigued in how people are appropriating current VMC technologies to go beyond "talking heads" conversations (which can be awkward to sustain for prolonged periods of time), to shared experiences that range from everyday routine events to major life events

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 73–90, 2013.

[6,18,20,25]. Prior work included stories of awkwardly chasing a toddler with a laptop to share with grandparents or snaking a webcam into an engine compartment to diagnose a problem with a father. This need for appropriation identifies ways that current technologies do not directly support the use of video to enable sharing an activity (other than talking) together over distance. We wanted to explore how to design systems that mitigate the issues inherent in VMC (i.e. background noise, privacy, framing – see [18,25]) and foster doing an activity together with someone in a different location, i.e. to remotely share an experience.

We focus on watching video together across distance as a specific kind of remote shared experience. Watching movies, TV shows, or other online video content is a popular social activity. Yet, with the dispersion of friends and family across different geographic locations, users might be remote from the people with whom they would like to watch together. We wanted to explore how adding watching videos together to VMC can use video to naturally sustain sharing activities together. This approach contrasts with adding communication to enable social TV watching over distance, which has been extensively explored in prior work [7,9,24,26]. We use watching video together not only as a remote shared experience to design for, but as a specific activity to learn more about remote shared experiences in general. One particular question we wanted to investigate was whether the combination of audiovisual entertainment with audiovisual communication would result in a divided-attention task in which neither activity was pleasant or effective.

In this paper, we describe a three-phase study that explores how people use VMC (e.g., Skype, Google+ Hangouts, FaceTime) in tandem with remotely watching videos together to create a shared experience. We present the concepts behind shared experiences and the relevant literature that has explored this area within Human Computer Interaction. We describe our three-phase methodology that integrates a survey, exploratory field study, and comparative lab experiment. We analyze the results of our study to provide guidelines on how to design systems to better support video communication during remote shared activities.

2 Related Work

2.1 Beyond Talking Heads to Shared Spaces

Recent work shows that VMC has multiple benefits when the field of view shifts from "talking heads" and towards visually sharing an environment [19,22,27]. Research has shown that shared visual environments can improve communication [22,27], task performance [22] and help establish common ground [8]. Common ground is the mutual awareness between a group of people in understanding the state of the task, as well as another member's state, intention, and comprehension in relation to the task [8]. By achieving common ground, group members can be more focused on completing the task as opposed to coordinating each other's roles in regards to it.

2.2 Shared Experiences

While much work has explored sharing visual environments for collaborative work, a recent wave of work has explored ways in which shared spaces and VMC can create shared experiences in home and consumer settings. Previous work has shown a desire for geographically separated family and friends to stay in touch and be part of the day-to-day activities [6,18,20,25]. Unlike the work focused on understanding how VMC improves collaborative tasks, this wave of work is focused on how it can help maintain and improve the social relationships between people. We expand this work by focusing on watching videos together as a shared activity domain.

2.3 Leisure Activities and Social TV

In our research, we focus on how pairs of remote individuals share leisure activities. Previous work has shown that regular participation in leisure activities has positive effects on physical and mental health [13] and dyadic relationships [2]. In this study, we focus on people who are remotely watching video programs together -- watching video programs (e.g., broadcast TV, recorded movies) together but from different locations. We chose this activity because it is a good example of a leisure activity that can easily be done remotely. When synchronized, the video program can act as a shared space and create a common ground for conversation.

In recent years, many researchers have built systems that embed a communication channel into the TV console (i.e. SocialTV or iTV) [7,9,24,26]. While much of the research has focused on building systems that share audio or text messaging while remotely watching TV together, some research has focused on the social interaction in these settings [11,12,15,23]. Dezfuli *et al.* found that people most often gathered together locally to watch video programs with family but had a high preference for watching with close friends [11]. Ducheneaut *et al.* found that the social discussion among the people was largely shaped and carefully crafted to fit in with the flow of the TV program, both when watching collocated and when watching remotely with an audio connection [12]. Harboe *et al.* ran field studies to see if people enjoy using audio-based SocialTV, how they behaved and what problems they faced doing this activity [15]. They found that people valued the experience, even if reluctant to try it at first. Metcalf *et al.* found that people felt constrained by pre-set text messages and wanted richer forms of communication while watching video programs [24].

We expand on Dezfuli *et al.* and Duchenaut *et al.'s* work on watching video programs locally by testing their findings on relationship strength and conversational patterns on people watching video programs remotely. We use the insights found from Metcalf *et al.* and expand on Harboe *et al.*'s work by running a field study that uses *video* chat as the means of communication. While Harboe *et al.* speculate that having video chat compete with the TV program for visual attention will be problematic in SocialTV, no research has been done to validate this speculation. Unlike Harboe *et al.*, we predict that people will enjoy video chat while watching video programs.

Our research focuses on enhancing the experience of video communication by augmenting it with the leisure activity of watching videos together. This approach complements most of the work on SocialTV that enhances the leisure activity of TV watching by augmenting it with communication. We build on Flora and Segrin's exploration on the effect of different leisure activities on relationships [14], by focusing specifically on watching video programs together and better understanding how to support it. While Aaltomen *et al.* [1] used groups of strangers to study how VMC configuration affected the experience of a remote shared activity, we use groups of people with an established relationship and who enjoy spending time together.

3 Methodology

In this section we describe the three studies comprising this paper. We combine our findings of the three phases and present the results at the end based on key themes we identified that were consistent throughout the phases.

3.1 Initial Survey

To contextualize our studies, we wanted to understand the relevance of watching video together in people's everyday lives. In particular, we had two questions in mind:

- Are people watching video programs together remotely or interested in doing so?
- If people are participating in this activity, how are they doing it?

We began exploring the answers to these questions with a survey. We ran this survey across two groups: full time employees and interns in our global software engineering company. A random selection of 1000 full time employees was invited to do this study via e-mail. All interns working during the summer received an invitation through a distribution e-mail list. 106 participants (59 full time, 47 intern) completed the survey. The majority of participants were male (78% in the full time group, 64% in the intern group). The majority of full time employees were in their 30's (49%) and 40's (32%). Almost all intern students were in their 20's (94%).

The survey consisted of 37 questions separated into five sections: basic demographics, current practices, desired practices, desired interaction, and miscellaneous. In basic demographics we collected age, gender, video program watching habits and video chat habits. Most importantly, we asked if they had ever communicated with someone at a different location while watching the same video program. This determined which sections they needed to complete. The number of responses to any question varied due to this branching and because all questions were optional.

Participants who had tried to watch video programs remotely completed the current practices section. It explored what they watched, with whom, what devices they used to watch video programs, and what devices and applications they used to communicate. Participants who had attempted or wanted to watch video programs remotely completed the desired practices and desired interaction sections. The desired practices section was almost identical to current practices, allowing us to compare what participants wanted versus what they did. In desired interaction, participants ranked which forms of interaction (e.g., touch, sight, talking) were most desirable as well as the best configuration to see and hear someone from a different location. All participants completed the miscellaneous section. Here, participants were able to share their thoughts and experiences regarding watching video programs remotely. The survey took approximately ten to fifteen minutes to complete.

3.2 Field Study

While the survey was effective in gauging how many people were interested in or regularly doing this activity and how they *thought* they did it, we were also interested in seeing how people *actually* did this and what elements resulted in high and low levels of social presence. We ran a field study to explore these questions.

For the field study, we selected 56 participants (29 male, 27 female) from an in-house database of usability subjects. 59% of participants were in their 20's, 25% were in their 30's, 9% were in their 40's, none were in their 50's, and 7% were 60 and above. Participants signed up for this study in pairs. Each participant knew their partner previously and was in a different location from them during the study. 13 of the pairs were same-gender pairs (7 male, 6 female), while the remaining 15 pairs were mixed-gender. 36% of pairs were in romantic relationships, 32% were close friends, 29% were immediate family and 7% were acquaintances. 96% of participants reported using video chat applications (e.g., Skype, FaceTime, Google+ Hangouts) at least once or twice a year – 27% using them on a weekly basis.

Participants scheduled a time with their remote companion to watch a video program (e.g., broadcast TV, DVD, internet streaming) together. 15 minutes prior to watching the program, participants logged into Skype and initiated a group video chat with their remote companion. For an additional gratuity credit, participants had the option of adding our auto-answer Skype account to the group video call so we could record it. Once the group video chat was established, participants started the video program at the same time. While the program was running, participants were encouraged to act as they normally would while watching video programs together. After the video program was complete, participants disconnected the video call and completed an online survey. Participants were rewarded for their participation with credit that could be traded for company merchandise. Figure 1 shows a typical set up pairs used to complete this study.

Fig. 1. A field study pair; both watched a program on TV and used video chat on their laptops

We studied the quality of their shared experience, the role of the communication device and the issues and events that affected the quality of their experience. To measure the quality of their shared experience, we looked at social presence, enjoyment and closeness. We used Biocca et al.'s definition of social presence, described as "a sense of co-presence with a mediated person and an awareness for their psychological, emotional and intentional state" [3]. We selected nine questions from the Networked Minds Social Presence Scale [4] most relevant to the nature of our study. Each question was asked on a 7-point Likert scale[1]. We also asked participants to rate how close they felt with their remote companion after completing the activity on a 7-point Likert scale. This construct was added since Kirk et al.'s work described closeness as the main reason behind people's usage of domestic video chat [20]. Finally, we asked them to rate how much they enjoyed the activity on a 7-point Likert scale.

We explored the role of the communication device in various ways. First, we investigated if the communication device was different or the same as the device streaming video programs. Second, we looked at the placement of the device. Last, we measured how the setup affected the video program watching experience in terms of connectedness and interference. We measured connectedness by asking participants to rate how much video chat made them feel connected to their remote companion during the activity. We measured interference by asking participants to rate how much video chat interfered with their experience of watching TV. These measures were in the post-task survey but also examined in the video data as well.

To understand the issues that affected each pair's shared experience, we analyzed it from a technical and context-specific perspective. Technically, we collected audio and video quality ratings of the video chat on 7-point scale plus a rationale behind each rating via open-ended responses within the post-task survey. We also looked at the video data for any technical issues that arose. Contextually, we qualitatively analyzed the video data based on the relationship between a pair of participants, where their attention was focused, and the type of interaction that occurred between each pair.

3.3 Lab Study

The do-it-yourself nature of the field study and its between-subjects design introduced significant variability that made it difficult to compare the different device configurations. We ran a within-subjects lab study that focused on comparing the differences between watching video programs in the same room, watching video programs and using video chat on the same device, and watching video programs and using video chat on different devices.

Eight pairs of participants were recruited from the in-house usability database. They were screened for watching video programs and using video chat on a fairly regular basis, and enjoying video programs with others. All were from the Puget Sound area (6 male, 10 female) ranging from 20 to 60 years of age (median 31).

After being briefed on the study and giving their consent, participants chose a 60 minute television program that they would be interested in watching during the study.

[1] All 7-point scales are reported with 1 = most negative, 4 = neutral, and 7 = most positive.

We had five prerecorded shows, which included drama, comedy, and a sporting event. Pairs tried three conditions of watching together – two being remotely connected using video chat. After each condition, they individually completed a short survey; because all questions were optional the number of responses to any given question varied. At the end of the study, they participated in a post-task interview. The study took approximately two hours to complete and participants were rewarded for their participation with credit that could be exchanged for company merchandise.

There were three experimental conditions. Our baseline condition, *Local*, had participants watching a third of the program side by side (Figure 2A). In the *Picture-in-Picture (PIP)* condition they watched a third of the program in separate rooms. The remote person's video was overlaid in the corner of the TV screen and their audio came from a speaker placed near the TV (Figure 2B). Our final condition, *Proxy*, was like PIP but the remote person's video and audio came from a separate device placed near them (Figure 2C). Condition ordering was counterbalanced. In the remote conditions we used high-quality speakers, high-definition webcams, and lavalier microphones to mitigate audio crosstalk.

Fig. 2. Lab study conditions: (A) Local – watching in the same room; (B) Picture-in-Picture (PIP) – watching with remote person appearing in inset on TV, which is highlighted here with an orange rectangle; (C) Proxy – watching with remote person on separate device

We reused our *shared experience* measures from the field study (social presence, enjoyment and closeness) and integrated them into our post-condition surveys. We did this to compare the experiences in the field and lab studies for any overlapping trends. In addition to the quality of shared experience, we expanded on the concept of *enjoyment* and revisited *connectedness* by asking participants to state which condition they enjoyed most/least as well as which condition they felt most connected with their companion. We asked participants to describe the differences between watching video programs remotely and locally. We specifically asked them to reflect on how the second screen's location in the Proxy condition affected their experience. We also asked if they knew anyone with whom they would like to do this activity with in the future.

4 Results and Discussion

Together, our three studies demonstrate that remotely watching TV together is a desirable, fun activity that augments social communication and reinforces social connection. We found that having the communication channels on a separate device was associated with better outcomes than picture-in-picture on the TV. We found that communication media richness corresponded to better outcomes. Finally we found that there are many technical challenges that can substantially degrade the experience. We present the results of the three studies together to focus on the insights supported by these converging lines of evidence.

4.1 Emerging Practice

Watching video programs, such as television, movies and online content, is often a social activity: 30% of respondents to our initial survey reported watching video programming with one or more other people every day; 75% reported doing it daily or weekly (n=105). At the same time, a substantial amount of watching is done solo: 40% reported watching video programs alone every day; 72% reported doing it daily or weekly (n=95, Figure 3).

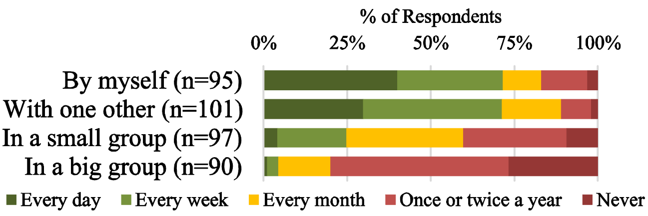


Fig. 3. Responses to the initial survey question, "*When watching video media, I watch it...*"

Some of our respondents (24%) had already communicated with someone in a different location while watching the same video program at least once, using a variety of communication channels, while others (26%) expressed interest in trying it (n=105). Significantly more young people (<30 years old) reported having watched video programs together remotely than older people (39%, n=49 vs. 11%, n=56, Mann-Whitney U=987, *p*=.001, Figure 4).

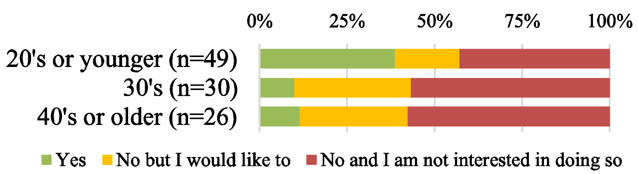


Fig. 4. Responses to the initial survey question, *"Have you ever communicated with someone at a different location while watching the same video program?"* by age bracket

While many of our respondents wanted to watch together remotely, 50% of our initial survey respondents reported being disinterested in communicating with someone at a different location while watching the same video program (n=105). When exploring reasons behind this disinterest, two consistent themes emerged. Some people are simply not interested in watching video programs:

> *"TV isn't a primary focus of my activities."* [Survey P7][2]

Others reported that they do not like communicating verbally while watching:

> *"In general I don't find watching TV together as a greatest social activity. People either watch and don't speak with each other, or speak with each other but then don't watch. With that I don't get the point of doing it together online…"* [Survey P39]

Prior research also found that some participants find watching video programs as time alone to unwind from work [15]. Other participants found certain genres like news and films to require more conscious attention on the video program and allowed less attention for socializing [11].

Following our studies, most participants responded positively when asked if they would be interested in watching video programs with a remote partner again (field study: 88%, n=56; lab study: 88%, n=16). This interest is substantially higher than the combination of the 24% who had tried it and 26% interested in trying it in our initial survey. This difference suggests that getting people over the hump of first trying it may increase the potential audience for this kind of remote shared activity. Harboe *et al.*'s found similar results with sharing audio chat while watching TV, although our results demonstrate video chat as a desirable form of SocialTV [15].

Insight: Some people, particularly young adults, are creating their own remote video watching experiences. There is an opportunity for bringing this capability to a larger audience, especially once they have had the opportunity to try it.

4.2 Enjoyment

Most participants (80%) in our field study responded positively when asked if they enjoyed the experience (n=55, M=5.4, SD=1.5). Several field study participants also commented on their enjoyment using Skype while watching together:

> *"It is fun because you can comment and talk about things that are happening. [My study partner] and I watched a lot of tv together in college and it really reminded me of that experience."* [Field P15a]

Likewise, most participants in our lab study had enjoyable experiences in the remote conditions with 81% responding positively (i.e. *somewhat agree*, *agree*, or *strongly agree*) in enjoyment (n=31, M=5.6, SD=1.3). A significant main effect of condition (Local, PIP and Proxy) was found on enjoyment (Friedman, X^2=17.92, p<.001). Posthoc pairwise comparisons revealed no significant difference between the Local and Proxy condition (Z=-1.996, p=.046) but the PIP condition was rated significantly lower than both the Local and the Proxy conditions (Z=-3.204, p=.001 and

[2] Participants are identified by study phase (Survey, Field or Lab), participant number, and a designator (a or b) if they were part of a pair.

Z=-2.705, *p*=.007 respectively). During the post-study interview, some participants described watching video programs remotely as a fun way to spend time together:

> *"Sometimes you can't be together but want to have Survivor parties or something. It's a great way to hang out if you actually couldn't."* [Lab P7a]

A few participants reported that using Skype while watching the same video program was initially awkward:

> *"It was a new experience and slightly awkward at first but then realizing we were laughing at the same things was fun"* [Field P17b]

Several participants reported, unsurprisingly, that it was not as pleasant as watching together in person.

> *"I still was able to easily communicate with my companion, as well as see their instant reactions to exactly what I was watching. However, because I was not physically in the room with them, it didn't seem to be as enjoyable."* [Field P13b]

Previous research provides possible reasons for this preference. Gaze and gestures are difficult to interpret during video chat, making an assessment of body language difficult [17]. Participants in video chat have a harder time reading their companion's mood and attention which can hinder effective interaction and negatively affect the shared experience [14]. Similarly, video chat lacks a shared physical space. Some people simply feel that digital representations of their companion are insufficient in comparison to having them physically there [16]. While we did not expect it to be as good as watching together in the same room, our goal was to explore enabling a shared watching video experience when they could not be together physically.

Some participants felt that the communication enhanced their enjoyment of the content.

> *"It was far more enjoyable than watching it alone; it's nice to discuss the show or make comments!"* [Field P27a]

Others felt that the communication lessened their enjoyment of the content.

> *"I'm more distracted from the video program and can't fully enjoy it."* [Field P17a]

The field study questionnaire results also showed mixed results regarding whether communication interfered with the content: 31% agreed, 40% neutral, and 29% disagreed (n=55, *M*=4.0, *SD*=1.6).

Insight: Most people enjoy the experience of remotely watching video, but challenges remain to make it more natural and to reduce the conflict between the video program and the communication.

4.3 Social Presence and Connectedness

We used nine questions from Biocca & Harms' Social Presence Inventory [4], along with our additional question of *closeness*, to gauge participants' level of social presence in the field study (Figure 5).

SP1. I often felt as if my remote companion(s) and I were in the same room together.
SP2. My remote companion(s) were often aware of me.
SP3. My remote companion(s) paid close attention to me.
SP4. I was easily distracted from my remote companion(s) when other things were going on. (Note: Response scale was inverted for analysis and presentation.)
SP5. I was sometimes influenced by my remote companion(s)' moods.
SP6. My remote companion(s) were able to communicate their intentions to me.
SP7. My thoughts were clear to my remote companion(s).
SP8. My actions were often dependent on my remote companion(s)' actions.
SP9. The behavior of my remote companion(s) was often in direct response to my behavior.
- - Watching video media together remotely made me feel closer to my remote companion(s).

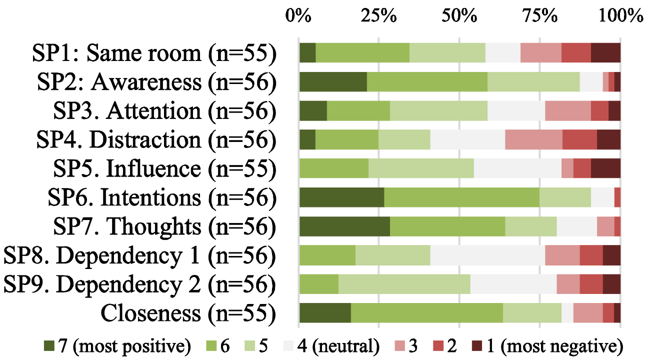


Fig. 5. Responses to the field study post-test survey questions relating to social presence

In order to compute a composite social presence score, we utilized a principal components factor analysis to extract relevant factors from these ten questions. Two different factors (factor load greater than .5) were revealed. The first factor included six of the questions from the Social Presence Inventory (SP1, SP2, SP3, SP4, SP6, and SP7) along with the closeness question. The second factor included the remaining three questions from the Social Presence Inventory (SP5, SP8, and SP9).

The first factor was significantly (p<.01) correlated with users' feelings of connectedness (r=.75), their enjoyment (r=.60), and their willingness to do the activity again (r=.52). As a result, we used the mean of these seven measures to compute the composite "Social Presence" (*SP*) score (the resulting score was in the range of 1-7, with 1 being low and 7 being high). The second factor was not significantly correlated with any of the other measures (r<.1) so these questions were not used in subsequent analyses.

In the field study, 86% of participants' SP scores were positive (>4) (M=5.09, SD=.97). Examining all of the demographic data for the participants (gender, age, level of experience with Skype, relationship with their partner), as well as their equipment configuration (device used for program and communication, and placement of the device) revealed no significant impact on the resulting SP scores (using Kruskal-Wallis test with alpha=.01). This may be a result of high individual differences across these variables, as well as the fact that most participants enjoyed their experience, regardless of the setup.

We used the same measure of SP in the lab study (Figure 6). For the two remote conditions, participants' enjoyment ratings were significantly correlated with their SP scores (PIP: r=.74, p<.001; and Proxy: r=.54, p=.027). This correlation supports previous work that found individual satisfaction (or enjoyment) from a leisure activity was the most influential factor on relationship satisfaction [2]. A significant main effect of condition (Local, PIP and Proxy) was found on SP scores (Friedman, X^2=30.63, p<.001). Posthoc pairwise comparisons revealed that Local had significantly higher SP scores, than Proxy (Z=-3.367, p=.001) or PIP (Z=-3.724, p=.001), and Proxy had significantly higher SP scores than PIP (Z=-2.976, p=.003). These results support prior literature comparing VMC to face-to-face communication [16,17]. They are also consistent with Media Richness Theory [10], suggesting that the communication media fidelity plays a strong role in the social connection of the experience.

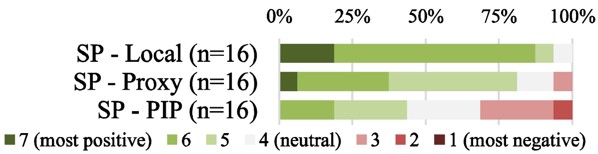


Fig. 6. Social Presence scores from the lab study post-condition survey. (Note: SP scores have been rounded to nearest integer.)

Most participants expressed that using VMC while watching video programs made them feel connected with their remote partner. In the field study, 75% of the participants responded positively that the experience made them feel connected (n=56, M=5.3, SD=1.3). Qualitative data from the field study provides some reasons why:

> *"Being able to hear comments from my friend helped me feel connected with him."* [Field P4a]

> *"It made me feel like I was sharing the same emotion with him...and we laughed etc together!"* [Field P8b]

> *"My sister lives two states away, it felt as if we were sitting next to each other watching a show like the old times."* [Field P16a]

Other comments suggested that the reason for feeling connected could be the common ground provided by the shared activity.

> *"You get to enjoy something together, rather than just talk."* [Field P2b]

> *"just skyping makes me feel as i need to be talking about something when i dont really have something to say but when watching media we can comment on whats going on"* [Field P12b]

> *"Well, when you are doing the same thing while Skyping, you have that little bit more in common with them. It definitely makes it more satisfying."* [Field P24a]

These results matched our hypothesis that the video program could serve as a shared space and coincides with Clarke and Brennan's concept of grounding [8]. The shared reference enabled participants to spend more time enjoying the activity and discussing the show or personal topics rather than spending the time trying to establish a shared perspective.

On the other hand, there is some design tension around competing video channels as some participants reported that watching the video programs distracted them from their normal Skype conversation behavior and decreased their sense of connectedness.

> *"When I usually Skype, I'm not doing anything else but Skyping, so it was a little strange because I felt like I was ignoring her... I think it takes away from the chat because you are watching the video therefore neither one gets your full attention"* [Field P11a]

> *"when I Skype with my husband it is usually to show him the baby or dog or important mail. Not eat up our time with a tv show"* [Field P25b]

O'Hara *et al.*'s work on mobile video chat [25] observed similar findings. Their participants described VMC as mentally demanding and, unlike mobile audio chat, made multitasking more difficult because the other party could see that you were not fully attending to them.

> **Insight:** For most people, remotely watching video resulted in a strong sense of connection or social presence. It strengthens ties when framed as an additional activity that increases the time spent together rather, than taking time away from talking together; controlling this framing is a significant design tension.

4.4 Placement of the Skype Window

In the initial survey, 48% of respondents ranked the picture-in-picture configuration most highly, where the video of their remote partner is shown embedded on the display showing the video program. 24% respondents wanted it on a separate device close to them, 16% wanted it on a separate device near the television, and 12% wanted it on a device they were holding (n=25).

In the field study, where users were able to choose whichever configuration they preferred or had access to, 61% used a picture-in-picture configuration and 39% used a Proxy configuration (n=54). Mann-Whitney U tests did not detect any significant differences between PIP and Proxy configurations in terms of enjoyment (Z=-.853, p=.394), SP (Z=-.116, p=.908) or their desire to do the activity again (Z=-.055, *p*=.956). The lab study however showed a clear benefit of Proxy over PIP (Wilcoxon Signed Ranks Test) for both participants' level of enjoyment (*Z*=-2.705, *p*=.007), and SP scores (*Z*=-2.976, *p*=.003).

In the lab study post-test interview, eight participants reported liking the proxy in the position we had set, five wanted it at the same distance but more in line of sight of the television, and three wanted it farther from them and closer to the TV. Note that the Proxy could be positioned to more easily see both the video content and the remote person than the head turning needed if sitting side-by-side in the same room.

It should be noted that the PIP conditions of the initial survey and lab study are not identical to that of the field study. In the former, PIP referred to both content and communication being placed on a television screen viewed from a distance. In the latter, all PIP configurations devised by our participants were on a laptop or desktop viewed close up.

Media Richness Theory [9] suggests that the Proxy condition would be preferred over PIP. The Proxy condition presented the remote companion on a larger screen, allowing for a higher quality representation of the individual. Likewise, the separation of TV and Proxy speakers allowed the participant to directly perceive the sound sources as spatially distinct. The visual and audio representations came closer to emulating the remote individual's physical representation and as a result made participants feel closer to having the actual person in the room [1,10].

Likewise, Grounding Theory [8] can further explain the Proxy's success over PIP. In the Proxy condition, a participant receives a visual cue when the other person shifts their visual attention between the content and the remote view. Because of this, less effort was needed to reorient between focused conversation and attending to the video program. For example, as seen in the video data, a participant could tell if their remote companion's attention was primarily focused on them based on how often they would turn their gaze toward the video chat screen. The participant in return could reciprocate and look back at the remote companion or keep looking at the TV, indicating that they are engrossed in the show and do not wish to chat right now.

Insight: When the video program is on a television, having the remote person on a proxy device on the coffee table is much preferred over picture-in-picture. Future work should compare these conditions to picture-in-picture on a laptop or tablet, as our results were inconclusive.

4.5 Communication Channels

Remote social TV can employ a variety of communication modalities. Prior work has focused on textual [7,9,11,24,26] or audio [12,15] modalities. The use of VMC in this application is largely an unexplored design space. Respondents to our initial survey ranked which modalities they felt were most important to help them feel connected. Talking verbally (i.e. audio) was rated highest, followed by text, then visual, then touch (Figure 7).

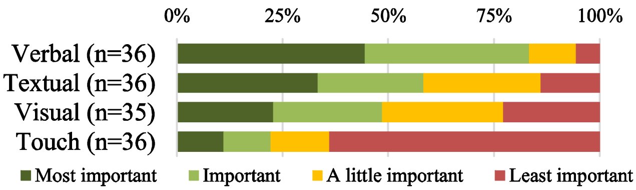


Fig. 7. Responses to the initial survey question, *"What interaction is needed to feel most connected with the person you are watching TV remotely with?"*

As noted above (Figure 6), in the lab study the Proxy condition had more positive outcomes than PIP. The most notable difference between these conditions is the greater visual angle subtended by the remote video. This result suggests that the video of the remote partner adds a key element to the experience.

While Figure 7 shows that touch was the lowest-ranked medium, 10% of respondents ranked it #1 and 10% ranked it #2 (n=36). In the field study questionnaire we asked, *"If there was a magical device that could increase your feeling of being connected with your companion(s), what key function or property does it need to have?"* Several respondents described wanting some form of touch:

> *"For me it's spatial. I want to feel warmth from loved ones in close proximity... so something that could make me feel like she was sitting right next to me."* [Field P16b]

Many researchers have investigated haptic interpersonal communication (e.g. [5]) and watching video together remotely may be a scenario that is ripe for exploring it.

Insight: While audio is fundamental for communication, video appears to play a key role in creating social presence. Tactile communication deserves more study.

4.6 Technical Problems

There were many technical hurdles that our field study participants needed to surmount, including synchronizing the content, handling the multiple audio streams, and general computer/networking glitches. Pairs spent a median of 3.6 minutes on setup (n=24). Twelve pairs continued to have technical problems throughout the session. We discarded the data of one pair because they were unable to solve the technical issues. 28% rated their audio experience negatively (n=54, M=4.6, SD=1.5); and 22% rated video negatively (n=55, M=5.1, SD=1.5).

These ratings were significantly correlated with participants' enjoyment (audio: r=.40, p=.003; video: r=.36, p=.008), SP (audio: r=.454, p=.001; video: r=.315, p=.019), and willingness to do the activity again (audio: r=.356, p=.008; video: r=.275, p=.042). This supports previous work that lists technical components as one of three factors (along with social and psychological) that affect the experience of communication [1].

While Liu *et al.* have developed a system for synchronizing video in a popular IM system [23], there is currently no widely-available standalone mechanism for perfectly synchronizing video playback between a pair of computers in different locations. For prerecorded programming, most participants used a technique where one counted down and then both pressed *Play*. Even broadcast programming is not perfectly synchronized since different broadcasting pipelines (e.g., satellite, cable, antenna) introduce different delays. While sync issues can be less noticeable in text chat, the slightest sync issue is much more obvious in video chat. Eight pairs in the field study had sync issues. Four pairs mitigated the sync issues by muting the content audio at one side and listened to it via Skype audio. When the content was out of sync even slightly, it impeded the interaction between the participants.

> *"My companion became annoyed that I was trying to make it so we were watching the shows at the exact same time. But if we weren't it was strange because we would laugh in different places."* [Field P18b]

These findings replicate similar findings from studies that suggest minor sync differences within video chat have significant negative effects on communication [21,22].

Audio was problematic because the microphone, which was intended to pick up the viewers' speech, also picks up the video program's soundtrack. While the automatic echo cancellation capabilities of the computer can keep the speech signal clean, it typically could not suppress the echo of the video program audio. When this occurred, a participant would hear the program audio both directly and via the communication channel, adversely affecting the engagement:

> *"Imagine sitting in a room watching TV, and having another TV in the same room, same show, but the timing is off a little bit and the volume is turned up."* [Field P19b]

Even without audio crosstalk, the two audio streams competed. Speech could drown out the program content, or vice versa.

> *"it was hard to hear the other person when they laughed or what not."* [Field P25b]

Insight: Technical complexity, content synchronization, and audio crosstalk are barriers to remote video watching. Technical problems can decrease the effectiveness of the overall experience.

5 Conclusion and Future Work

We found that remotely watching video programs together while connected by VMC is engaging, fun, and fosters social bonds between the participants. Our study documents the important factors that enable this particular "shared experience" [6] to lift VMC "beyond talking heads."

We found little to suggest that the combination of audiovisual entertainment and audiovisual communication caused divided attention. Audio levels did compete, but having two separate video streams did not. Duchenaut et al. [12] noted that the pauses and redundant content in TV shows create opportunities for communication and may mitigate this competition. The Proxy configuration's advantage over PIP may have been in part because it helps people better manage their attention. We believe that there is further opportunity for better understanding and designing for attention in shared-experience systems.

We believe that a key affordance of VMC is how it enabled more immediate and natural sharing of reactions, which is an important part of enjoying watching videos together. Video, compared to text or only audio, is well-suited for richly sharing laughter and other emotional reactions to the video program. The users' comments suggested that this sense of sharing these reactions through video contributed to the positive scores in social presence, connectedness, and enjoyment.

We believe that practical systems for remotely watching video together would have positive social impact by allowing close ties to strengthen their bonds by sharing experiences despite physical distance in a way that is different from, and complementary to, simply conversing together. Our next step is to develop software – and possibly hardware – to support watching together remotely. There are many interesting design and technical problems – initiating the experience, choosing the program to watch, closely synchronizing playback, and solving audio crosstalk.

Finally, watching TV is but one of many possible remote shared experiences. This study strongly supports rich media beyond audio communication in remote shared experiences. This is a rich design space that deserves more exploration.

References

1. Aaltonen, V., Takatalo, J., Hakkinen, J., Lehtonen, M., Nyman, G., Schrader, M.: Measuring Mediated Communication Experience. In: QoMEx 2009, pp. 104–109. IEEE (2009)
2. Berg, E.C., Trost, M., Schneider, I.E., Allison, M.T.: Dyadic Exploration of the Relationship of Leisure Satisfaction, Leisure Time, and Gender to Relationship Satisfaction. Leisure Sciences: An Interdisciplinary Journal 23(1), 35–46 (2001)
3. Biocca, F., Harris, C., Burgon, J.: Toward a more robust theory and measure of social presence: review and suggested criteria. Presence: Teleoperators and Virtual Environments 12, 456–480 (2003)
4. Biocca, F., Harris, C.: Networked Minds Social Presence Inventory (Scales only, ver. 1.2) (2003), `http://cogprints.org/6742/1/2002_netminds_scales.pdf` (accessed September 19, 2012)
5. Brave, S., Ishii, H., Dahley, A.: Tangible interfaces for remote collaboration and communication. In: CSCW 1998, pp. 169–178. ACM Press (1998)
6. Brubaker, J., Venolia, G., Tang, J.: Focusing on Shared Experiences: Moving beyond the camera in video communication. In: DIS 2012, pp. 96–105. ACM Press (2012)
7. Cesar, P., Chorianopolous, K., Jensen, J.J.: Social Television and User Interaction. ACM Comput. Entertain. 6(1) (2008)
8. Clark, H.H., Brennan, S.E.: Grounding in communication. In: Resnick, L.B., Levine, R.M., Teasley, S.D. (eds.) Perspectives on Socially Shared Cognition, pp. 127–149. APA, Washington, DC (1991)
9. Coppens, T., Trappeniers, L., Godon, M.: AmigoTV: Towards a social TV experience. In: Proc. EuroTV 2004, vol. 36 (2004)

10. Daft, R.L., Lengel, R.H.: Organizational information requirements, media richness and structural design. Management Science 32, 554–571 (1986)
11. Dezfuli, N., Khalilbeigi, M., Mühlhäuser, M., Geerts, D.: A Study on Interpersonal Relationships for Social Interactive Television. In: Proc. EuroITV 2011, pp. 21–24. ACM Press (2011)
12. Ducheneaut, N., Moore, R.J., Oehlberg, L., Thornton, J.D., Nickell, E.: Social TV: Designing for distributed, sociable television viewing. International Journal of Human-Computer Interaction 24(2), 136–154 (2008)
13. Fitzpatrick, T.R.: The Quality of Dyadic Relationships, Leisure Activites and Health Among Older Women. Health Care for Women International 30(12), 1073–1092 (2009)
14. Flora, J., Segrin, C.: Joint Leisure Time in Friend and Romantic Relationships: The Role of Activity Type, Social Skills and Positivity. Journal of Social and Personal Relationships 15(5), 711–718 (1998)
15. Harboe, G., Massey, N., Metcalf, C., Wheatley, D., Romano, G.: The Uses of Social Television. Computers in Entertainment 6(1), article 8 (2008)
16. Heath, C., Luff, P.: Disembodied conduct: Communication through video in a multimedia environment. In: CHI 1991, pp. 99–103. ACM Press (1991)
17. Hirsh, S., Sellen, A., Brokopp, N.: Why HP People Do and Don't Use Videoconferencing Systems (2005)
18. Inkpen, K., Taylor, B., Tang, J., Junozovic, S., Venolia, G.: Experiences2Go: Sharing Kid's Activities Outside the Home with Remote Family Members. In: CSCW 2013. ACM Press (2013)
19. Junozovic, S., Inkpen Quinn, K., Blank, T., Gupta, A.: IllumiShare: Sharing any surface. In: CHI 2012, pp. 1919–1928. ACM Press (2012)
20. Kirk, D., Sellen, A., Cao, X.: Home Video Communication: Mediating 'Closeness'. In: CSCW 2010, pp. 135–144. ACM Press (2010)
21. Krauss, R., Bricker, P.D.: Effects of transmission delay on the efficiency of verbal communication. Journal of Acoustical Society of America 41(2), 286–292 (1967)
22. Kraut, R.E., Gergle, D., Fussell, S.R.: The Use of Visual Information in Shared Visual Spaces: Informing the development of virtual co-presence. In: CSCW 2002. ACM Press (2002)
23. Liu, Y., Shafton, P., Shamma, D.A., Yang, J.: Zync: the design of synchronized video sharing. In: Proc. DUX 2007, article no. 12. ACM Press (2007)
24. Metcalf, C., Harboe, G., Tullio, J., et al.: Examining presence and lightweight messaging in a social television experience. ACM Trans. Multimed. Comput. Comm. Appli. 4(4) (2008)
25. O'Hara, K., Black, A., Lipson, M.: Everyday practices with mobile video telephony. In: CHI 2006, pp. 871–880. ACM Press (2006)
26. Shirazi, A.S., Robs, M., Schleicher, R., Kratz, S., Müller, A., Schmidt, A.: Real-time nonverbal opinion sharing through mobile phones during sports events. In: CHI 2011, pp. 307–310. ACM Press (2011)
27. Whittaker, S.: Things to talk about when talking about things. Human-Computer Interaction 18, 149–170 (2003)

Exploring Reactions to Widespread Energy Monitoring

James A. Colley[1], Benjamin Bedwell[2], Andy Crabtree[1], and Tom Rodden[1]

[1] Mixed Reality Lab, School of Computer Science,
University of Nottingham, Nottingham, UK, NG8 1BB
[2] Horizon Digital Economy Research
University of Nottingham, Nottingham, NG7 2TU
{psxjac,first.last}@nottingham.ac.uk

Abstract. This paper explores the measurement, apportionment and representation of widespread energy monitoring. We explicate the accountability to users of the data collected by this type of monitoring when it is presented to them as a single daylong picture. We developed a technology probe that combines energy measurement from the home, workplace and the journeys that connect these spaces. Through deployment of this probe with five users for one month we find that measurement need not be seamless for it to be accountable; that apportionment is key to making consumption for communal spaces accountable and that people can readily make useful inferences about their energy consumption from daylong pictures formed from widespread monitoring. Finally, we present four issues raised by the probe – the nature of real world monitoring, the dynamic and social nature of apportionment, disclosure of energy data and alignment of incentives with consumption – that need to be addressed in future research.

Keywords: Distributed energy monitoring, measurement, apportionment, representation, technology probe.

1 Introduction

Energy security and sustainability are established as global societal challenges and have prompted a variety of responses and initiatives. Technical and operational responses have turned to digital technologies, to capture the consumption of energy and to make this available to stakeholders to optimise energy use. The ability of the technology that captures and makes available information on energy use has been a major focus for researchers in HCI and ubiquitous computing. Researchers have sought to promote awareness of energy consumption as a way of encouraging changes in peoples' behaviour [7]. A critical feature of much of this work has been the development and deployment of devices and interfaces that surface energy consumption in order to make consumption accountable to consumers and thus motivate them to reduce energy use. Examples include technologies for the home [13, 14], transport [10] and the workplace [23].

The framing of the energy challenge within HCI in terms of surfacing consumption in particular settings has not passed without criticism. Pierce and Paulos [22], for

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 91–108, 2013.

example, have called for a shift beyond residential energy feedback systems, highlighting a disconnect between current HCI energy research with its emphasis on in home display; and energy research that adopts a much broader perspective. Others have questioned the emphasis on measurement and persuasion as a narrowing of the concerns involved in sustainability [2]. However, considerable efforts persist in making energy measurement even more pervasive and fine-grained. For example, research is emerging within ubiquitous computing that focuses on capturing a rich picture of energy use across a broad range of settings [e.g, 10,21]. As these technologies for energy data capture mature, a number of researchers have offered the vision of a personal energy monitor that captures peoples' energy footprint throughout the day [16].

We are interested in understanding how people may react to a future world where information about their energy use throughout the day, across the multiple spaces they operate in, is captured and surfaced. What does it mean to surface energy consumption and make it accountable across and between the different places that people inhabit and how will people react to this? By way of an initial answer, this paper presents the development and study of a technology probe focusing on the measurement and surfacing of energy consumption across and between multiple spaces.

The probe combines a mobile device carried by users that works in tandem with energy monitoring systems in different spaces. The probe monitors consumption in users' homes and workspaces, and records the time spent in these places along with journeys made by users. Three key research issues drove the design of the probe:

1. **Measurement**. How *widespread* and how *fine-grained* does measurement need to be to surface energy consumption across and between multiple spaces?
2. **Apportionment**. How should measured energy consumption be apportioned across and between *multiple spaces*?
3. **Representation**. Will representations of apportionment make energy consumption across and between spaces accountable to consumers, and in what ways?

The probe was deployed in the wild with five participants over a period of one month. At the end of the month users were presented with a web-based representation of their data displaying consumption across their home and work lives as well as the journeys that are taken. The data was used to drive semi-structured interviews with each participant in order to understand the representation and its potential affordance to human reasoning about energy consumption.

The results from our probe suggest that while it is not possible to measure all of a person's daily consumption, partial measurement alongside user inference is sufficient to make the consumption accountable to users. However, that ability to make energy information available across different contexts raises a number of significant tensions that may cause widespread measurement and individual accountability to act as a significant source of demotivation for some. This is particularly true when we consider the differences between the roles of a user in domestic and workplace energy consumption.

2 Existing Approaches towards Surfacing Consumption

Recent research in HCI has focused on surfacing energy consumption to encourage consumers to change their behaviour. The basic premise is that surfacing energy use makes consumption visible and accountable to consumers and, in turn, enables them to take appropriate action. To date, there are two key approaches towards surfacing consumption: space-centric approaches and person-centric approaches.

2.1 Surfacing Energy Consumption within Spaces

A common approach has been to surface energy consumption within a *space*. This often relies on measuring the consumption of appliances, and results in the creation of artefacts or persuasive displays that surface appliance-based energy consumption data and imply accountability. Thus far, these feedback systems have focused mainly on domestic spaces. Examples include the power aware cord [13], the power socket [17], the watt bot [19], light bulb radiators [14] and the minimalist in-home energy consumption display [24]. While novelty effects have been seen, significant long-term changes in behaviour have not yet been demonstrated by these HCI-led initiatives. However, long term studies of standard in home displays (IHDs) have demonstrated an impact. For example, in a review of 36 studies of feedback on behaviour between 1995 and 2010, Ehrhardt-Martinez et al. found an average of 9.2% reduction in electricity usage as a result of the use of an IHD [8].

Work within ubiquitous computing has focused on low-cost sensing technologies to capture more detailed information about domestic consumption. Patel et al. [21] have developed a system that makes use of electrical noise signatures on residential power circuits to detect the switching of particular appliances. This demonstrates how very specific high-resolution sensing can allow micro-level events in the home to be surfaced. Abrahamse et al. [1] reviewed thirty-eight studies around technologies to capture and present energy consumption. It is notable that none surfaced consumption from more than one space and that the predominant focus was on surfacing energy use in the home. An additional review of studies of 'eco-feedback' by Froehlich et al. [10] echoes the concern that studies of such technologies seldom consider the connected nature of consumption in everyday life and that people are therefore made accountable for consumption in their homes in isolation from other spaces. As the majority of consumption still occurs *outside* the home[1], we are interested in exploring how measurement and feedback technologies might include, but also go beyond domestic infrastructures to consider energy use across *multiple* spaces.

2.2 Surfacing Personal Energy Consumption

Efforts to surface consumption within a space have been complemented by other efforts to surface *personal* consumption. The UbiGreen application [10] makes use of mobile phones connected to Mobile Sensing Platform devices to semi-automatically

[1] E.g., 71% of energy is consumed outside of the home in the UK [6].

infer consumers' transport modes, exploiting various visualisations to then make carbon consumption accountable to the user to promote greener travel choices. Related work by Mun et al. makes use of map matching and hidden Markov models to build a "Personal Environmental Impact Report", which visualizes the carbon intensity of trips. This work is technically impressive in the way that it combines multiple data sources however there is no data on its users' reactions [20]. Surfacing personal energy use has also become a key feature in government initiatives. These have promoted concepts such as the carbon footprint and Personal Carbon Trading Schemes [e.g., 9]. The underpinning idea is that individuals are made accountable for their carbon footprint by having to deduct their personal portion of consumption away from an allocated quota based on an overall national cap.

An alternative person-oriented approach seeks to understand the human practices that shape energy consumption and sustainable behaviour [5]. This has highlighted the broad character of personal energy consumption.

"Traditionally, we have taken a within household approach to infrastructure, emphasizing systems that will make our homes smart. But households are connected to a variety of infrastructures beyond the home … householders desire insight into … resource usage and more visible real-time information on within the home resource consumption. They also desire more information on between-homes consumption" (ibid.)

Chetty and Grinter's study draws attention to the distributed nature of energy consumption and the concomitant implication that personal energy consumption needs to bridge the home and other spaces that people inhabit if it is to be of broad utility. This view is supported by research that has explored the development of the 'personal energy meter' [16]. This trades on a vision of a widespread energy-monitoring infrastructure to provide people with a personal account of energy use across their lives.

3 The Technology Probe

Existing approaches to surfacing energy consumption and making it accountable to users suggest the need to move beyond isolated contexts to surface energy use across multiple spaces. While there is a broad turn towards capturing detailed information about energy consumption in more locations through the widespread introduction of smart meters, the distributed character of energy consumption raises a host of challenges. How might we surface consumption across the multiple spaces we inhabit? How might we make consumption at work accountable to users? How might we understand energy use in public spaces that are seldom monitored?

Capturing information is one thing, making it into a useful and usable resource for action another. Making use of existing monitoring infrastructures, the aim of this paper is to contribute to an increased understanding of how people might respond when energy consumption from multiple spaces is surfaced and presented to them. We address the challenge through the development and deployment of a technology probe. Technology probes are an extension of the cultural probe methodology [12], which seeks to "provoke inspirational responses" from potential users as a resource for design. Technology probes on the other hand are "thought-provoking technologies" – provoking in the sense that they enable people to reason about how future visions "do or don't fit into their lives" [18]. They also have the added benefit that they enable data logging. While involving design and implementation, technology probes are not intended to be prototype solutions amenable to usability testing.

"They are not changed during the use period based on user feedback. In fact, a deliberate lack of certain functionality might be chosen in an effort to provoke the users" (ibid.)

Our interest is not in whether the technical intervention is usable, but rather in the principles underpinning future widespread monitoring technologies. The probe seeks to elicit views on capture and attribution and how we might design choices of energy apportionment. This is reflected in a number of explicit design choices. Most notably we have chosen a simplistic means of apportionment to provoke users to comment on their own personal accountability for energy use. We do not seek to record and represent the fact that different inhabitants will consume energy in a space in different ways. Likewise, we do not consider that individuals may stake claim to energy consumed in spaces that they do not inhabit (e.g., a dishwasher timed to commence while its owner is at work). We are interested in the strategies of apportionment that make sense to users and how these design choices might be reasoned about. Thus, rather than build a variety of apportionment models into the probe, we used a deliberately crude but consistent approach as starting point, which we could then open up to discussion during interview.

The probe itself combines three key elements designed through a series of iterative refinements: multi-site energy measurement, energy apportionment, and representation of personal energy consumption.

We describe each of these in turn below, before moving on to consider the results of deployment and users' reactions to the probe.

3.1 Measurement

Monitoring centred on a cloud-based logging service that received data from three distinct types of data source: home energy consumption collected via a deployed meter; workplace energy consumption collected by interfacing to a building management system, and the reported points of journeys taken, collected using a smartphone.

Monitoring the Home

The home monitoring kit consisting of a current transformer monitoring clamp and a plug computer was used to record the rate of consumption in participant's homes. Data was sent to a secure cloud-based server every 6 seconds. On the server side, power readings were aggregated once every hour down to a resolution of 1 mean reading for every 5 minutes to reduce the computational expense of querying the data.

Monitoring the Workspace

Energy consumption in the various workplace buildings in our study was logged by an estate management department using an existing building management system (BMS). Automated BMS export functions were used to deliver data to the probe's cloud-based logger. The BMS monitors are coarse at a temporal level, as they are designed to generate data for high-level monthly reports, providing readings per building every 30 minutes. Furthermore, each building's readings vary in resolution: for some, lighting and socket power readings per floor are recorded and for others only the consumption of the whole building is recorded. Newer buildings tend to have a higher resolution of monitoring compared to older ones, where retrofitting more monitors is deemed not cost-effective.

Monitoring the Journey

Location information is collected through a dedicated mobile phone application, which reports the users' location based on Wi-Fi positioning via Google's location service on the Android platform. Wi-Fi was chosen both because it provides a level of accuracy suitable to determine which building a user currently inhabits (between 6 and 20 meters on average) and because it works inside and outside of buildings, while not draining the battery significantly. The service typically provides location updates (consisting of latitude-longitude coordinates) every 30 seconds. The location is sent securely along with a unique identifier for the user to the cloud-based logger.

3.2 Apportionment

The apportionment process collates the data relevant to a particular individual for a particular day based on a set of monitored spaces. Monitored spaces are defined using 'geo-fencing', i.e. the definition and storing of polygons that represent the boundaries of workplace buildings and homes inside which energy consumption is being monitored. These polygons are cross-referenced with users' locations as they are reported by the logging application on their mobile phones to determine whether an individual is present in any of the monitored areas. If entry into an area is detected (i.e., the user's mobile device automatically 'checks-in') this event is logged; no further latitude-longitude coordinates are logged until the user's mobile leaves ('checks-out' of) the area.

This record of automated check-ins and check-outs provides a simple means of apportioning energy consumption data across multiple spaces. It is based on the crude assumption that every individual inhabitant of a monitored space is responsible for an equal proportion of the energy used in that space while they are present. As all inhabitants of a space were not monitored, individuals were made accountable for personal energy consumption equal to the total energy consumption of the space divided by the maximum occupancy of that space. This is obviously a crude presumption – it will allow high consumers to "free-ride" – and as such we would not necessarily propose it for a fully deployed system.

Discovering an appropriate model of apportionment has already gathered some significant research interest. For example, Hay et al. have investigated different models of apportionment for the workplace using access logs and energy monitoring to simulate policies for example individuals [15]. Our research interest focuses on the issues of intelligibility and accountability of apportionment. As such, a crude model that is simple to understand but potentially unfair allows us to takes a provocative stance to encourage reaction from users.

3.3 Representation

The way that data was represented to our participants was key in provoking them into thinking about the issues implicated in widespread energy monitoring, particularly the accountability of energy consumption when measured and apportioned across multiple spaces. We represented the data gathered during the deployment of the probe via a web-based ‘personal energy dashboard’, which was made available to participants during the interviews to surface and replay their energy usage and encourage them to reflect on their consumption patterns.

The Personal Energy Dashboard

The personal energy dashboard (Figure 1) consists of 4 components:

1. A map (Figure 1, Section 1) showing the trail that the user has left when they are in unmonitored space, the spatial boundaries of monitored spaces they have occupied, and consumption in the places they have occupied. The consumption is represented on the map by bars situated at the centre of the associated building with a height proportional to the energy consumption of the building for the time the user has occupied it. This map assists users in thinking about how they connect consumption infrastructures to their lives.

A day long bar chart (Figure 1, Section 2) with time on the horizontal axis and consumption on the vertical axis. This shows how consumption has varied as the user moved through different buildings throughout the day. No bar is shown if the

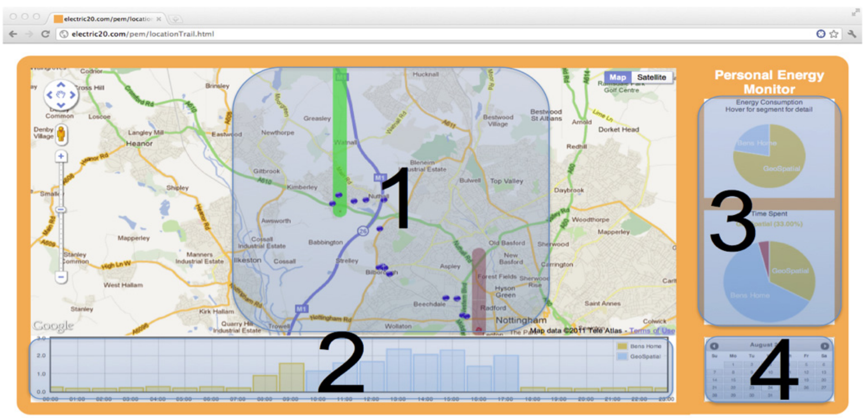


Fig. 1. The personal energy dashboard

user has been in unmonitored space for an hour slot. The purpose of this chart is to show how consumption changes throughout the day depending on the spaces inhabited. It also demonstrates the apportionment done by the system so as to allow users to reflect upon and challenge it.
2. Two proportional pie charts (Figure 1, Section 3). One shows the proportion of consumption for the time spent in monitored space (Section 3, top). The other shows the segments of time spent in monitored buildings; this also includes a segment for time spent in unmonitored space (Section 3, bottom). This feature is intended to provoke views around apportionment and consumption across multiple infrastructures.
3. A date selector (Section 4, Figure 1), which allows the user to revisit previous days by navigating through the log on a day-by-day basis.

The display is intended as a focal point for our semi-structured interviews rather than as an interface to promote energy reduction. The aim is thus reflection on the underlying approach and systems rather than actionable energy reduction suggestions. The personal energy dashboard allows users to both reflect upon the time they spend within different monitored and unmonitored spaces throughout their day, and to view the consumption data that pertains to them within those spaces.

4 The Probe Study

The probe deployment involved five participants recruited via email within one workplace that was chosen because researchers could readily access data from the workplace's BMS. We installed monitoring equipment in each participant's home and provided a mobile phone that reported his or her location to our system. The participants included:

- George, a 27 year-old professional who is recently married. He often spends weeks living alone as his wife travels internationally on business. He works in an office from 9 to 5 most days and occasionally from home.
- Alice, a 31 year-old, married, middle-class office worker. She shares use of a car with her husband, but enjoys cycling to work when her husband needs the car. Alice works long hours in the office, but her work can occasionally spill into her home life.
- Cecile, a 34 year-old single woman from Europe. She generally works a 9-to-5 day in the office and has an active but distinctly habitual social life. She commutes by bicycle and is proud to consider herself both energy conscious and a low consumer of energy.
- Fred, a 32 year-old recently married professional who likes to keep his home and work life separate. He made the choice to live on the outskirts of the city where he works. He often drives to and from work but occasionally cycles when the weather is agreeable.
- Bill, 26 years old and currently in the final stages of his formal education. He commutes 150 miles (twice a week) by either car or train between his place of

study and his girlfriend's hometown, where he spends the weekend and works a part time job for 1 day a week.[2]

The probe deployment lasted one month, during which time a rich set of data about participants' energy consumption and their location was logged. Data from the probe was presented to users through the personal energy dashboard, allowing them to browse through their daily energy consumption over the period of the deployment. Semi-structured interviews were used to understand the accountable character of monitoring and apportionment across and between multiple spaces or, to put it another way, to explore how amenable this kind of information is to human reasoning.

5 Results

Energy data was collected from all 5 participants' homes and 4 separate workplace buildings used by participants. Cumulatively, ~183000 data points for energy consumption were recorded across all monitored spaces, and ~4500 location reports outside of monitored space were recorded. This level of measurement allowed us to create energy dashboards that represented a month's worth of data for each participant.

The first and most obvious thing we noticed is that our participants spent a considerable proportion of their time in unmonitored space (11%). This included commutes to and from work, time out socializing, visits to the country and holidays. Their remaining time was split between personal space (62%) and the work place (27%). There were obviously individual differences in this data reflecting different lifestyles. For example, Fred who lives furthest from his place of work had a larger proportion of time in unmonitored space, reflecting his daily commute. Similarly, Cecile had a large proportion of unmonitored space reflecting her regular attendance at social events. There were also significant differences between the working week and weekends, where our participants often spent 50% of their time in unmonitored spaces.

While our participants spent the majority of their time in personal space, most energy consumption actually took place in communal spaces. Over 63% was consumed in communal spaces where our individuals had limited control of their consumption, with less than 37% being consumed in personal space where they had direct control of consumption (see Figure 2).

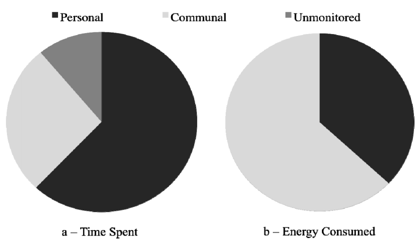


Fig. 2. Proportion of time (a) and energy consumed (b) in different spaces

[2] These are, of course, not the participants' real names.

The collected data provoked responses from our participants to the measurement, apportionment and representation of energy data and these three themes provided the focus for our interviews.

5.1 Measurement

The probe was designed to explore how widespread and how fine-grained measurement needs to be to surface energy consumption across and between multiple spaces. Deployment of the probe suggests that collecting sensor data in the wild, particularly from multiple sources, results in 'seams' [3] appearing within the data. These are a result of the environment (e.g., when users move outside of and between monitored spaces) and unexpected events (e.g., when users let their phones run out of battery or leave their phone at work overnight). When this seamful data was presented to participants through the dashboard interface they responded with discussions around what the seams meant to them and how they affected the legibility of the data.

> Alice: I can see there's no energy data here, but I can see where I was travelling to and from … I did lots of travelling this day but by different means. I took the train to Liverpool then drove back. I can see that from the routes.

Rather than being seen and treated as unfathomable gaps in the data, participants were instead able to see their behaviour in the seams and infer usage. Discussions around the location trails that were recorded outside of monitored space revealed that participants were particularly good at recalling their travel habits and filling in the gaps when they existed. Sometimes this was done at a level of detail that would be difficult to achieve by sensing and measurement alone:

> Fred: I took the train from there to there, then I cycled down [road x].
>
> Bill: That day I had to go a different way because of road works, it took me longer.

Participants demonstrated an understanding around issues of connectivity, how this impacted their data, and accepted that this is part of life.

> George: It didn't really get when I was out in the countryside, but there's no consumption there anyway. It's all fields and cows.

We designed the probe to allow participants to collect data from each other's houses when they visited them. Figure 1, Section 3 shows the spaces in which consumption occurred; when participants visited each other's houses they were able to see the visited house's consumption as part of their own chart. Our intent here was to surface and explore the issues of leakage that might possibly occur with monitoring of this form and to raise issues of privacy with the users.

On measurement, we conclude that it is meaningful for participants to have consumption in the home and work place measured, a load reading every hour taken from these spaces allowed them to compare and contrast the varying consumption between the spaces that they passed through and reason about their accountability for consumption in these spaces. We also note that gaps in data are acceptable as long as seams are meaningful enough to be repaired by user inference.

5.2 Apportionment

The probe was also designed to enable us to address the key question of apportionment and how it should be calculated across and between multiple spaces. Although we adopted a crude means of calculation, participants nevertheless expressed concerns over the level of control available over consumption in the home versus the workplace.

> Alice: I can't turn off my computer at work, because I need it to work. I can't turn the heating down even though its too warm here, at home I can just turn the heating off!

Participants were able to read from the dashboard that the energy monitored for their time at work far outweighed the energy consumption for their time at home.

> Bill: The work consumption makes my home look like a piss in the ocean.

However, people also elaborate rationales for these high loads in communal buildings, examples of which include recognizing that the energy consumption was necessary for health or a sense of shared responsibility for the provision of services to multiple users of buildings.

> Fred: I think I do feel some responsibility for all the servers on at work all the time but the servers at work mostly serve my building so they're related to our work.

When discussing apportionment of energy with participants, a desire for a more detailed breakdown of energy consumption in communal spaces was sought to assist in their reasoning about the high levels of consumption.

> George: I get that the consumption is high here, I mean the building is huge so it must cost a lot to run, but I'd feel more comfortable with this if I knew how much was used for heating, lighting, and other stuff. I can work that out in my home from my bill but I've got no idea here.

People were also sensitive to how the apportionment might affect the way they are viewed in comparison to others, particularly when a partial representation of carbon consumption was exposed.

> Fred: I've got so many friends who think because they recycle paper at home they're being green but then they drive 50 miles a day and they're probably in the top 5% of consumers in the world of energy.

On apportionment, we conclude that the way in which energy is apportioned to consumers, particularly when measured from a communal space is key in attempting to make it accountable. Our simplistic method of apportionment was sufficient for participants to understand their effect on widespread energy monitoring and afforded them sufficient reasoning to be able to make choices about how and where they work.

5.3 Representation

Our third key research question concerns the adequacy and appropriateness of representations of apportionment across and between multiple spaces – do they suffice to make the distributed character of energy consumption accountable to consumers, and in what ways? Our elaboration of the previous issues concerning the measurement and apportionment of widespread energy measurement for accountability already provide some insight into this; here we go into more detail.

Given that some of the recruits were friends and were likely to visit one another's homes, coupled with the sensitizing effect of coverage of location-based services in the media, it was anticipated that discussions surrounding data ownership and privacy would arise from use of the application. While not frequent, there were cases during the study where users visited one another, and in the interviews participants spoke about whom they would like to share data with and whom they would not:

> Cecile: I really don't like the idea of people I don't know being able to see my energy consumption. It's fair if I've invited someone to my house for them to see and share part of my energy consumption but I don't want random robbers seeing my consumption.

The issue here seemed to focus on control of disclosure and an associated view that friends had access to energy consumption as part of visiting the home.

Reflecting general concerns about the security of the Internet, participants raised issues around the sensitive nature of the data collected by the system. The concerns related to what might happen if people were to access the collected energy data and use it for ill gain. Sharing of location was viewed as less concerning than the sharing of data from the home. For the majority of participants in the study, there was an acceptance of location sharing as being part of modern every day life, with understanding around mobile applications that make use of locational data.

> Alice: I can tell you what I share my location with at the moment, [points to apps on phone] Movies, Facebook, Latitude, Foursquare, Voucher Cloud.

To this we would add that the data also reflected routine patterns of consumption and activity. The dashboard enabled participants to reason about routines and how these impacted their consumption across different spaces. Unsurprisingly, a regular pattern of consumption and space occupation was common in the data. People could readily see their activities in the data and provide accounts and rationales for the routines. Consider, for example, a week's of George's time charts that he explored in the interview using the lower chart in section 3 of the dashboard (Figure 1).

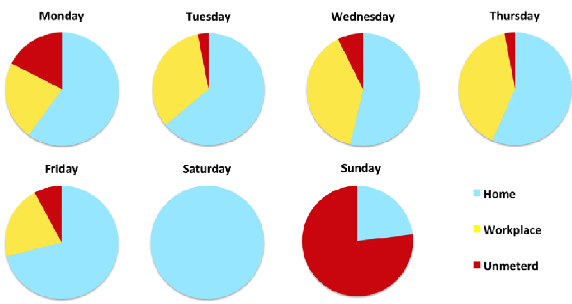


Fig. 3. A week's time apportionment charts for George

In this week George was at work each weekday (yellow), he also had some time in unmonitored space (red). On two days the time in unmonitored space increased as he was engaged in social activities. This was revealed during his interview:

> George: Ah, I left work early on that day, Monday. I was driving the wife to her piano lesson, the traffic was crap, was on the road for ages. I can see that 'cause that bit's bigger [points to unmonitored space section of chart].

On Saturday, George spent no time in communal or unmonitored space as a result of him staying at home to attend to household chores. George also paid particular attention to section 2 of the dashboard for weekend days. He noted that for the Saturday he spent at home the amount of energy consumed was higher than normal and he was able to attribute his actions to this high consumption

> George: Ha, that'll be me doing all the washing and drying then. Plenty of work done there.

The representations used made it possible for participants to identify their routines and associated consumption across multiple spaces they could easily spot significant changes to those routines and how this altered their consumption profiles.

When making the workplace consumption visible to participants through the dashboard, the accountability of that consumption to them as workers within the space was drawn into question.

> Bill: I can't be held accountable for the consumption here, even if it is high 'cause I'm only the doer of the work not the one ordering it to be done, or the one who controls how these buildings are run. I mean if it were up to me I'd just work at home and then none of it could be blamed on me

Here, Bill demonstrates how he feels that he should not be held to account for consumption in the workplace, as it is someone else's job to manage it.

The dashboard thus provided an interesting means of showing the impact of their everyday activities on their daily routines and the possibilities offered by widespread energy monitoring. Our results show that presenting consumption information for different buildings occupied by a user over a day affords strong accountability for the consumption that takes place inside these buildings, provided that some apportionment of communal consumption is applied.

6 Key Lessons and Future Challenges

As access to energy data grows, new opportunities for systems that surface consumption for people across multiple spaces arise. However, the reactions to our probe from individuals who were subject to the data that this kind of monitoring can provide suggest that there are a number of issues that the community must consider.

6.1 The Nature of Real World Monitoring

Our probe highlighted the messy nature of real world data collection. Ubiquitous computing often looks to the future outlining new services or presuming the existence of future capabilities. For example, the vision of a global personal energy meter described by Hay et al. [16] depicts fine grain modelling and the ability to show how much energy has been used as a result of our daily activities at the level of the appliances used and activities conducted.

Our probe sought to explore the extent to which this vision may currently be realised in the real world. Our probe does not offer fine-grained measurement. The deployed infrastructure did not exploit appliance-level monitoring in the home or fine- grained monitoring in communal spaces. Rather we sought to exploit systems in widespread use or in the process of widespread deployment. Working with existing sensing systems is a key feature of real world monitoring and is likely to be a perennial feature for any energy monitoring system given the nature of the built environment where legacy plays a critical role.

Energy consumption data is often monitored for reasons other than apportionment to individual people (e.g. to support the maintenance of buildings). This means that unambiguous attribution of measurement is likely to remain technically very demanding. Gaps and seams in data were ever-present in this data collection of the probe and we would suggest this a perennial feature of this class of system.

Rather than seeing these gaps as problematic we would suggest an approach drawing upon ‘seamful design’ [3] to exploit unmeasured space and limited accuracy. Letting people see the gaps in data allows them to make inferences about their activities, provoking thoughtful interactions with energy. In fact, the ease with which participants could infer or recall the activities that generated their data suggests that there may not be much need to process the data further to reveal patterns: self-reporting may instead be a useful (and enjoyable) means of a user improving the accuracy of their own data.

6.2 The Dynamic and Social Nature of Apportionment

The probe made use of a very simple apportionment model to elicit reactions from people to the notion of apportionment for energy, particularly in communal spaces. Building management systems in communal buildings often represent consumption on a much greater scale and with less granularity than the home. This can make it difficult for individuals to relate to data from communal spaces. However, this is not the only reason for the lack of a personal view of accountability. The complex nature of our interactions with energy throughout our lives means that the principles of apportionment are highly contextualized in practice. Consequently, negotiation regarding the cause of energy use is integral to understanding energy consumption and attributing responsibility. In the home, apportionment is negotiated amongst a small, intimate social network. In larger communal spaces this network of accountability can be very broad, including people and practices unfamiliar to the individual, dissuading the individual from taking responsibility.

Ubiquitous computing systems that seek to encourage some level of personal accountability within communal spaces should consider how to leverage the broader social networks in communal spaces. Who is responsible for consumption and how much of the consumption for a space they are responsible for will change constantly depending upon the members of the space and roles that they take. This is not a simple issue. We suggest that the development of this kind of system should consider the need for a human interface to the apportionment models applied, allowing all actors in a system to have influence over the way in which they are made accountable for consumption.

6.3 Disclosing Energy Data

Participants' responses to the probe highlight the sensitive nature of energy data and awareness that energy use could tell others a great deal about an individual's everyday activity and routine. The idea of sharing energy data with other people for various reasons was accepted by our participants: the participants appeared to have an understanding that they were often donating personal data as they used their phones, computers, and other technologies, but that they tended to receive some kind of reward for doing so (improved services, financial offers, etc.). Participants' reactions to the probe indicated that it was not the inference that someone might draw from their data that was of concern, but that there appeared to be no clear mechanisms to prevent others from accessing their data without permission.

Participants drew on a range of analogies such as inviting people into the privacy of their home to suggest suitable mechanisms that they were familiar with. This highlighted how the nature of relationship greatly affects the control a user might expect over the disclosure of their data. For example, when considering disclosing energy data to a commercial partner, participants spoke of formal contracts, laying out specific terms and returns. When considering disclosure to friends and family, participants related to practices of visiting suggesting with respect to the data from their home that if they were happy to allow someone through their front door then they were happy to share their energy data with them.

This leads us to suggest that personal energy monitoring systems should provide mechanisms for controlling disclosure of personal energy data and that these mechanisms should be modelled on existing metaphors for, or be integrated with existing practices of, sharing and disclosure. These mechanisms should allow for renegotiation as and when necessary, and explicitly involve a discussion of the value returned to the individual in disclosing the data.

6.4 The Alignment of Incentives with Consumption

People are currently exposed to multiple incentives to reduce energy consumption. In the workplace, the motivation is often provided in terms of carbon reduction as part of initiatives that aim to improve the public image of a company or organization and reduce running costs. In the home people are exposed to raw data in the form of meters and tariffs, as this is a space in which people directly incur cost for their consumption. This direct relationship suggests the use of tariffs and individual monetary reward as an incentive for change.

Our probe also surfaced energy consumption from different spaces that users inhabit as they go about their everyday lives. This invited cross-comparison between these spaces and promoted user reflection on the amount of energy used in each context. Understanding and managing the transitions between these spaces and the possibility of linking and comparing energy use across them will become a critical issue to consider as energy monitoring becomes increasingly widespread.

Offering individuals greater access to their energy consumption opens up the possibility of foregrounding the different incentives and drivers towards energy reduction. In doing so, this raises issues of the extent to which these are complementary or in conflict with each other. There is an as yet unexploited opportunity to transfer good practice through common mechanisms for surfacing energy use and unifying incentives, making people accountable for their consumption across all aspects of their lives in a consistent manner, and rewarding them for managing their throughout the day.

7 Conclusion

We have presented the deployment of a technology probe used to elicit user reactions to the measurement, apportionment and representation of energy consumption across multiple spaces. Our probe combined energy monitoring systems from the home and the workplace allowing us to study reactions by users towards the measurement and apportion of energy use. This has involved exploiting representations of personal energy consumption to investigate accountability of consumption measured across the home and the workspace.

A striking feature of our probe was the extent to which the measurement of energy consumption intertwines with a broad range of other social, political and economic drivers. The measurement and display of energy consumption surface these in terms of issues around apportionment and accountability. The future design of any widespread energy monitoring system will, by the very nature of the energy data collected, be embedded within a range of concerns about the nature of sustainability [7].

Our probe suggests that these issues will play out in the ways in which the technology seeks to attribute energy use and the ways in which the presentation of consumption provides information about the different contexts of energy use. Most striking for us was the ways in which energy consumption at work dominated total energy use but was not matched by the ability for people to exercise control over this use.

The probe highlights the importance of understanding and aligning with the broad social, political and economic drivers at play. Measurement and presentation of energy use is unlikely to change users' behaviour without some form of initiative to encourage and reward change. It is important that we understand the initiatives and that our measurement aligns with them rather than undermines them.

Our final observation is the way in which any ubiquitous computing system designed to measure energy embodies a sense of accountability. The challenge here is the way in which accountability and attribution of use are encoded in the system. Energy apportionment and accountability is a dynamically negotiated process. For example, the apportionment of energy use from a utility can often be a source of major discussion and debate involving consideration, compromise and trading. People need to understand the rationale for apportionment at play and may often wish to negotiate the principles by which this apportionment is carried out.

References

1. Abrahamse, W., Steg, L., Vlek, C., Rothengatter, T.: A review of intervention studies aimed at household energy conservation. Journal of Environmental Psychology 25, 273–291 (2005)
2. Brynjarsdottir, H., Håkansson, M., Pierce, J., Baumer, E., DiSalvo, C., Sengers, P.: Sustainably unpersuaded: how persuasion narrows our vision of sustainability. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 947–956. ACM, New York (2012)
3. Chalmers, M.: Seamful Design and Ubicomp Infrastructure. In: Proceedings of Ubicomp 2003 Workshop at the Crossroads: The Interaction of HCI and Systems Issues in Ubicomp (2003)
4. Chetty, M., Brush, A., Meyers, B.: It's Not Easy Being Green: Understanding Home Computer Power Management. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1033–1042. ACM, New York (2009)
5. Chetty, M., Tran, D., Grinter, R.E.: Getting to Green: Understanding Resource Consumption in the Home. In: Proceedings of the 10th International Conference on Ubiquitous Computing, pp. 242–251. ACM, New York (2008)
6. Department for Energy and Climate Change. UK Climate Change Sustainable Development Indicator: 2010 Greenhouse Gas Emissions, Provisional Figures. Technical report, Department for Energy and Climate Change (2011)
7. Disalvo, C., Sengers, P., Brynjarsdóttir, H.: Mapping the landscape of sustainable HCI. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1975–1984. ACM, New York (2010)
8. Ehrhardt-Martinez, K., Donnelly, K.A., Laitner, J.A.: Advanced metering initiatives and residential feedback programs: A meta-review for household electricity-saving opportunities. American Council for an Energy-Efficient Economy, Washington, D.C. (2010)
9. Environmental Audit Committee. Personal Carbon Trading. Fifth Report of Session 2007-2008. House of Commons (2008), `http://www.publications.parliament.uk/pa/cm200708/cmselect/cmenvaud/565/565.pdf`
10. Froehlich, J., Dillahunt, T., Klasnja, P., Mankoff, S.C., Harrison, B., Landay, J.A.: UbiGreen: Investigating a Mobile Tool for Tracking and Supporting Green Transportation Habits. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1043–1052. ACM, New York (2009)
11. Froehlich, J., Findlater, L., Landay, J.: The design of eco-feedback technology. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1999–2008. ACM, New York (2010)
12. Gaver, W., Dunne, A.: Projected Realities Conceptual Design for Cultural Effect. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 600–607. ACM, New York (1999)
13. Gustafsson, A., Gyllenswärd, M.: The Power-Aware Cord: Energy Awareness through Ambient Information Display. In: CHI 2005 Extended Abstracts on Human Factors in Computing Systems, pp. 1423–1426. ACM, New York (2005)
14. Gyllensward, M., Gustafsson, A., Bang, M.: Visualizing Energy Consumption of Radiators. In: IJsselsteijn, W.A., de Kort, Y.A.W., Midden, C., Eggen, B., van den Hoven, E. (eds.) PERSUASIVE 2006. LNCS, vol. 3962, pp. 167–170. Springer, Heidelberg (2006)

15. Hay, S., Rice, A.: The case for apportionment. In: Proceedings of the First ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings, pp. 13–18. ACM, New York (2009)
16. Hay, S., Rice, A.: A global personal energy meter. Adjunct Proceedings of the 7th International Conference on Pervasive Computing (2009)
17. Heller, F., Borchers, J.: PowerSocket: towards on-outlet power consumption visualization. In: CHI 2011 Extended Abstracts on Human Factors in Computing Systems, pp. 1981–1986. ACM, New York (2011)
18. Hutchinson, H., Hansen, H., Roussel, N., Eiderbäck, B., Mackay, W., Westerlund, B., Bederson, B.B., Druin, A., Plaisant, C., Beaudouin-Lafon, M., Conversy, S., Evans, H.: Technology Probes: Inspiring Design for and with Families. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 17–24. ACM, New York (2003)
19. Jönsson, L., Broms, L., Katzeff, C.: Watt-Lite: energy statistics made tangible. In: Proceedings of the 8th ACM Conference on Designing Interactive Systems, pp. 240–243. ACM, New York (2010)
20. Mun, M., Reddy, S., Shilton, K., Yau, N., Burke, J., Estrin, D., Hansen, M., Howard, E., West, R., Boda, P.: PEIR, the personal environmental impact report, as a platform for participatory sensing systems research. In: Proceedings of the 7th International Conference on Mobile Systems, Applications, and Services, pp. 55–68. ACM, New York (2009)
21. Patel, S.N., Robertson, T., Kientz, J.A., Reynolds, M.S., Abowd, G.D.: At the flick of a switch: Detecting and classifying unique electrical events on the residential power line (Nominated for the best paper award). In: Krumm, J., Abowd, G.D., Seneviratne, A., Strang, T. (eds.) UbiComp 2007. LNCS, vol. 4717, pp. 271–288. Springer, Heidelberg (2007)
22. Pierce, J., Paulos, E.: Beyond energy monitors: interaction, energy, and emerging energy systems. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 665–674. ACM, New York (2012)
23. Schwartz, T., Betz, M., Ramirez, L.: Sustainable energy practices at work. In: Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries, pp. 452–462. ACM, New York (2010)
24. Yun, T.-J.: Investigating the impact of a minimalist in-home energy consumption display. In: CHI 2009 Extended Abstracts on Human Factors in Computing Systems, pp. 4417–4422. ACM, New York (2009)

HCI for City Farms: Design Challenges and Opportunities

Peter Lyle, Jaz Hee-jeong Choi, and Marcus Foth

Urban Informatics Research Lab, Queensland University of Technology,
130 Victoria Park Road, Kelvin Grove, Queensland, Australia
{p.lyle,h.choi,m.foth}@qut.edu.au

Abstract. Urban agriculture plays an important role in many facets of food security, health and sustainability. The city farm is one such manifestation of urban agriculture: it functions as a location centric social hub that supplies food, education, and opportunities for strengthening the diverse sociocultural fabrics of the local community. This paper presents the case of Northey Street City Farm in Brisbane, Australia as an opportunity space for design. The paper identifies four areas that present key challenges and opportunities for HCI design that support social sustainability of the city farm: A preference for face-to-face contact leads to inconsistencies in shared knowledge; a dependence on volunteers and very limited resources necessitates easily accessible interventions; other local urban agricultural activity needing greater visibility; and the vulnerability of the physical location to natural phenomenon, in this instance flooding, present a design challenge and a need to consider disaster management.

Keywords: Urban Agriculture, City Farm, Design, Sustainability, Urban Informatics.

1 Introduction

Over half the worlds population now resides in urban spaces, and this is an increasing trend. Urban agriculture is the use of urban and peri-urban spaces for the cultivation and production of food, fuel and livestock [1]. Urban agriculture contributes to the ability of cities to provide food to its inhabitants, and offers many positive benefits to society, in the areas of sustainability and health e.g. [2-5]. This domain has been identified as a space that could benefit from new types of HCI technology design [6], however the limited number of studies mean there is opportunity for exploring urban agriculture from different methodological approaches, including exploring the field as an 'opportunity space' [7].

This research presents the findings of a case study of the Northey Street City Farm in the city of Brisbane in Australia. A city farm is a term often used interchangeably with community garden, and is a form of urban agriculture that has a specific location

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 109–116, 2013.

and involves a productive farm or garden, run by the local community[1]. Fieldwork was undertaken onsite over six months in late 2011, and was followed by interviews in early 2012 to gain an understanding of the way the farm operates. Participant observation and semi-structured interviews generate data with which to reveal the challenges and opportunities to communication and resource management at the farm. This research forms part of a larger study exploring design opportunities and challenges within different manifestations of urban agriculture within Brisbane.

This approach attempts to provide a clearer picture for HCI designers who wish to engage with city farms. We identify four key outputs as a result of this research that relate to design. The opportunities and challenges described apply to the environments urban agriculture is practiced, far away from the 'standard office environment', as well as the more common limitations and problems that are experienced, designing for volunteer dependant organisations.

2 Prior Work

Urban agriculture as a subfield of sustainable HCI has previously been reviewed previously [8, 9], and the limited studies that explore links between urban agriculture and technology indicate opportunities for innovation to create greater community engagement [2, 6]. Urban agriculture was also a partial focus of a CHI workshop that explored design considerations for small scale agriculture and fisheries, focusing on economic and social sustainability [10].

Locally grown food generates fewer food miles than large scale rural agriculture where food must travel 1500-2500 miles before it is consumed [2], although this oft-quoted figure has been disputed, and the real value may be much higher [11]. Producing food locally in urban environments is important because it helps alleviate pressure on centralised rural food production to meet the food demands of cities [3]. Urban agriculture promotes public health, both with the availability of local fresh produce – an alternate to refined and processed foods, which contributes to obesity [4] – and provides physical exercise by engaging in the practice [5].

As a key related work Odom [2] explored urban agriculture in a similar setting to the focus of this work within Australia. Odom [2] by contrast, took the approach of ethnographic fieldwork over several months with two sites – a ground level garden, and a rooftop garden. A continuation of this is Odom [12], which also investigated different opportunities for design.

Agriculture and technology is also researched as part of ICT4D, although this is not specific to urban environments. Two examples of this in rural India [13, 14] explored methods of giving greater voice and connectivity to farmers.

3 Case Study: Northey Street City Farm

Northey Street City Farm (NSCF) is a non-profit community organisation and was Brisbane's first community garden, located within two kilometres on the northern side

[1] Federation of City Farms and Community Gardens,
`http://www.farmgarden.org.uk/farms-gardens`

of the CBD. It was started by a group of friends in 1992 who lived in the local area and was supported by the local council. It has since grown to become a non-profit organisation, which employs nine part-time staff, has a separate management committee that gives the farm direction, and a large group of volunteers.

NSCF was chosen for this study as it models several positive aspects that city farms strive for: The location provides fertile land for productive farming; it boasts facilities for practical education through volunteer participation and organised events; and, it offers space to host a local farmers market. NSCF has an established reputation that includes appearances in newspapers and on Australian television[2].

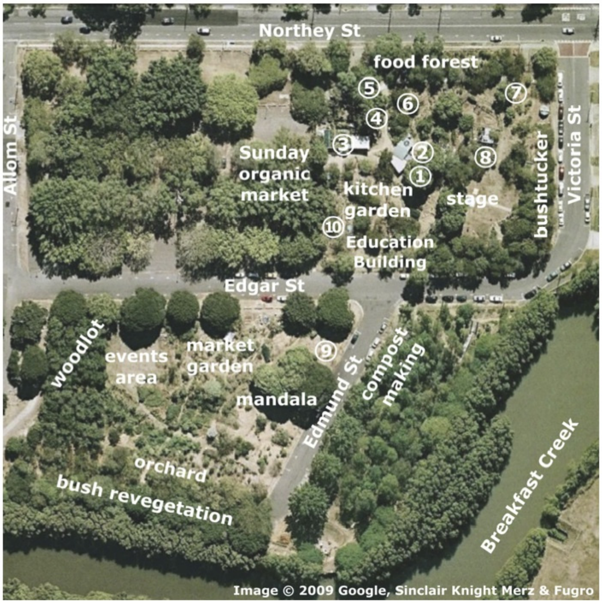


Fig. 1. Northey Street City Farm map provided as part of a guidebook to new volunteers

NSCF is financially supported through the following means: community grants, a weekly farmers market, and a plant nursery. These support a diverse range of activities at the farm, which are performed by paid staff and a variety of volunteers. The farm itself is comprised of several functional areas that can be seen on the sitemap in Figure 1 above. There is a paid staff member who is responsible for the management of each area. These include: a nursery, which provides income for the continued operation of the farm; a backyard garden to provide an example of how a typical backyard environment can be utilised for food production; a market garden, where food is grown to be sold at a weekly market, and; a kitchen garden, where food is grown and then prepared in the onsite kitchen, to be consumed by the staff and volunteers who participate at the farm. There are other smaller areas without dedicated staff, such as

[2] http://www.nscf.org.au/northey-street-city-farm-in-the-media/ Accessed March 25, 2013

the composting, native plants, an orchard and a mobile chicken coup, and a new area dedicated to allotment gardens.

At the time this research was conducted, NSCF was undergoing a review process to reflect on the policies and positions of the organisation. The primary author attended one of many meeting of volunteers and staff that discussed the review to resolve potential shifts/redistributions of staff roles and responsibilities.

4 Methodology

The fieldwork for this study was carried out over the course of three months. It utilised ethnographically oriented methods of participant observation and semi-structured interviews to embed the researcher within the urban agriculture context of NSCF. The rationale for this approach is to give a deep understanding of the current operation and communication channels that NSCF utilise to continue functioning. The purpose was to understand the way interaction and communication occurred between people at the farm, and to gain an appreciation of what potential challenges and opportunities HCI designers must consider. The data was then analysed to derive themes from the data, and the key findings are presented in the Discussion section below.

4.1 Participant Observation

After completing a 1-hour farm tour (that is run weekly as a means of introducing the farm and encouraging new potential volunteer participation), the primary author then volunteered once a week for two months, working with different groups at the farm. Following this period, the author made the occasional visits on Tuesdays, and the weekend food market on Sundays. This culminated in a number of informal interactions and discussions with staff and volunteers at the site. Observational notes were recorded after the days of participation.

4.2 Semi-structured Interviews

Participants for semi-structured interviews (20-40 minutes each) were recruited using a process of snowball sampling. This began with a paid member of the administrative staff, who helped with recruitment suggestions which resulted in five participants overall (a mix of paid staff and volunteers). These five interviews provided sufficient to gain an understanding of the different roles in the farm planning, management and operation, and the nature of communication at the farm. These semi-structured interviews took place onsite at the farm over a period of two days, for the purposes of determining the participant's involvement in the farm, their connection with other members of the farm community, and their connection with other urban agriculture communities. Participants had been involved with NSCF for a period ranging from six months to five years. Two of the participants were current representatives on the Management Committee (in addition to their role as a volunteer and a member of administrative staff), which overseas the general direction of the farm.

5 Discussion

Analysis of the interview and observation data provided insight into four key areas that could shape potential interventions by HCI designers. These include: a preference for face-to-face communication as a means to share information, often leading to confusion and mixed messages; the dependence of NSCF on the volunteer workforce, and the general resource limitation that requires any intervention to be easily accessible in terms of resource requirement; an opportunity to make other local urban agriculture practice more visible and accessible, while respecting the limited time volunteers and staff have to invest in their gardening passions; and, the need to acknowledge and account for the physical location of the city farm, which is likely to be located on land unsuitable for other high-value uses (e.g. residential or industrial) and in the case of NSCF, is on a floodplain.

5.1 Face-to-Face and Inconsistency

Face-to-face is recognised as the prime means of communication at NSCF, as all participants indicated. Face-to-face communication means that information is not always communicated to everyone for whom it may impact, in addition to other issues similar to those identified by Nardi and Whittaker [15]. NSCF has different levels of communication that match the hierarchy; there is a management committee, paid staff, team leaders, and a wider circle of people that consists of both paid staff and key volunteers, and then all other volunteers. An example of where the complexities of different groups, have led to inconsistencies is the management of the farm's online presence. The promotions manager indicated that she was confused as to whether she should be involved in reviewing content before it is made public via the website or Facebook page, as some staff members went directly to the web manager, and other areas of the farm setup their own Facebook pages that were managed independently. The number of other Facebook pages that the interview participants were aware of also differed.

Developing technology that can alleviate communication breakdowns, while respecting the preference for face-to-face interaction presents a unique challenge for HCI designers. This could come in the form of a prototype that enables recording of face-to-face transactions, possibly as audio recordings or automated transcriptions. Issues of cataloguing would then need to be considered, given resource limitations described in the next challenge.

5.2 Resource Limitations

NSCF is reliant on volunteers to continue operating, despite support from successful grant applications, the nursery, educational operations, the weekly market and annual membership costs. This is unsurprising given that the rise of urban land prices is often a factor in 'pushing' agriculture to the fringe and rural areas. Paid staff are expected to perform volunteer duties in addition to their paid hour allocation. While the space at NSCF is effectively utilised, there is insufficient land for it to operate as a viable commercial farm. Notwithstanding this, the primary focus at NSCF is education and community building, allowing people to learn and participate through volunteering.

In order for the farm to continue it's operation by paying staff, and procuring tools and materials, the farm understandably depends on incoming funds. A lack of surplus money and resources, and a demand to acquire the ability to continue operation means that resource allocation must be performed carefully and efficiently, as there is minimal margin of error. This is a common problem for both starting an urban agriculture project, as well as its continuation [16]. HCI designers should consider the overhead of any technologies, as organisations that depend on volunteers such as NSCF are unlikely to consider any investment in new systems. Utilising a participatory design approach may not succeed if the designers themselves are unable to see beyond what is simply said by participants, a problem identified by Bertelsen et al. [17]. Taking stock of existing infrastructure, or providing offsite infrastructure as part of any collaboration with a city farm, would have a greater chance of success.

5.3 The Bigger Picture of Local Urban Agriculture

All those interviewed expressed interest in other urban agriculture activities outside of NSCF, however only one interviewee indicated they had any time to actually participate. The reasons provided were related to the individual's preference to dedicate all of the volunteering time to a single initiative to encourage and engage with community based urban agriculture. Community based urban agriculture is distinct from that undertaken by individuals in their backyards, which is not necessarily dependent on interaction with others.

This limitation of time is not dissimilar to the Resource Limitation point discussed above, however the focus is on the individuals commitment. While participants did not think themselves able to invest time into multiple projects, they were all interested in knowing about other city farm, community gardens and similar community efforts to champion urban agriculture in the local area, providing an opportunity for technology to make this visible. An example might be to provide a geo-mapping experience that can show presents nearby city farms as a map overlay. Sharing expertise and knowledge between different urban agriculture projects, may provide a way to alleviate obstacles for newer initiatives, a problem identified by Kaufman and Bailkey [16].

5.4 Physical Environment

Urban agriculture tends to be forced out of urban spaces as populations increase, as the relative value of the land increases with higher population densities [18], so it is unsurprising that in order to be located so close to the CBD, it is located on flood plain.

Prior to the participant observation, at the end of January 2011 the farm suffered from flooding, causing damage to the farm that meant it was not operational for nearly four weeks. One participant in particular detailed the difficulty experienced keeping the farm operational during this time, and had to setup a remote working environment from her home. A similar problem occurred in January of 2013, with the farm once again flooded. After the floods email communication in addition to the regular farm newsletter was circulated to give an explanation of planned clean-up days, and advised of what facilities or utilities had been damaged (and included requests for donations or replacements). The limitations of the physical environment is consistent with Kaufmann and Bailkey's [16] experience that noted the physical setting of urban agriculture projects is subject to a number of issues and constraints, relating to the land (and possible soil contamination), as well as socially driven physical problems of security and vandalism.

As such the ability to manage and mobilise the community in times of natural disasters provides an opportunity for HCI designers to take advantage of mobile technology and telecommuting. This also highlights the importance of proper documentation and backup management of farm resources, perhaps taking advantage of cloud technologies such as Dropbox.

6 Conclusion

The role of city farms in the broader context of urban agriculture is important as it provides a central physical space for like minded people learn and participate in the process of growing food. HCI designers should consider the positive benefits from a thriving urban agriculture community such as the case of Northey Street City Farm, and take stock of the challenges and opportunities outlined in this paper in order to better share information internally, alleviate resource strains, make visible the bigger picture of local urban agriculture, and prepare for the physical limitations of the environment.

The findings of this paper coincide with that of Odom [12] and Kaufman and Bailkey [16], with regard to the potential value that could be added by improving the visibility of urban agriculture projects, not just to those not engaged, but also to those embedded within existing practice. As part of the larger project of which this study is a part, the future direction is to test the generalizability of the above findings with other manifestations of community (or other non-commercial) urban agricultural practice, such as that of grassroots movements (e.g. "guerrilla gardening", or the permablitz movement), and communities of practice who support each others back or front yard gardening endeavours.

Acknowledgements. This research study is supported under the Australian Research Council's Linkage Projects funding scheme (project number LP100100232).

References

1. Smit, J., Ratta, A., Bernstein, J.: Urban agriculture: an opportunity for environmentally sustainable development in sub-Saharan Africa. Environmentally Sustainable Division, Africa Technical Dept (AFTES), World Bank (1996)

2. Odom, W.: “Mate, we don’t need a chip to tell us the soil’s dry”: opportunities for designing interactive systems to support urban food production. In: Proceedings of the 8th ACM Conference on Designing Interactive Systems, pp. 232–235. ACM, Aarhus (2010)
3. Brown, K.H., Jameton, A.L.: Public Health Implications of Urban Agriculture. Journal of Public Health Policy 21, 20–39 (2000)
4. Chopra, M., Galbraith, S., Darnton-Hill, I.: A global response to a global problem: The epidemic of overnutrition. Bulletin of the World Health Organization 80, 952 (2002)
5. Howe, J., Wheeler, P.: Urban food growing: The experience of two UK cities. Sustainable Development 7, 13–24 (1999)
6. Blevis, E., Morse, S.C.: SUSTAINABLY OURS Food, Dude. Interactions 16, 58–62 (2009)
7. Hornecker, E., Halloran, J., Fitzpatrick, G., Weal, M., Millard, D., Michaelides, D., Cruickshank, D., De Roure, D.: UbiComp in opportunity spaces: challenges for participatory design, pp. 47–56. ACM (Year)
8. DiSalvo, C., Sengers, P., Brynjarsdóttir, H.: Mapping the landscape of sustainable HCI. In: Proceedings of the 28th International Conference on Human Factors in Computing Systems, pp. 1975–1984. ACM, Atlanta (2010)
9. DiSalvo, C., Sengers, P., Brynjarsdóttir, H.: Navigating the terrain of sustainable HCI. Interactions 17, 22–25 (2010)
10. Hirsch, T., Sengers, P., Blevis, E., Beckwith, R., Parikh, T.: Making food, producing sustainability. In: Proceedings of the 28th International Conference Extended Abstracts on Human Factors in Computing Systems, pp. 3147–3150. ACM, Atlanta (2010)
11. Schnell, S.M.: Food miles, local eating, and community supported agriculture: putting local food in its place. Agric. Hum. Values, 1–14 (2013)
12. Odom, W.: “You don’t have to be a gardener to do urban agriculture”: Understanding opportunities for designing interactive technologies to support urban food production. In: Choi, J.H.-J., Foth, M., Hearn, G.N. (eds.) Eat Cook Grow: Mixing Human-Computer Interactions with Human-Food Interactions. MIT Press, Cambridge (2013)
13. Rege, R.: Designing interactive information access technologies for small scale rural indian farmers. In: Proceedings of the 27th International Conference Extended Abstracts on Human Factors in Computing Systems, pp. 3359–3364. ACM, Boston (2009)
14. Patel, N., Chittamuru, D., Jain, A., Dave, P., Parikh, T.S.: Avaaj Otalo: a field study of an interactive voice forum for small farmers in rural India. In: Proceedings of the 28th International Conference on Human Factors in Computing Systems, pp. 733–742. ACM, Atlanta (2010)
15. Nardi, B.A., Whittaker, S.: The place of face-to-face communication in distributed work. Distributed Work, 83–110 (2002)
16. Kaufman, J., Bailkey, M.: The Lincoln Institute of Land Policy: Farming inside cities: entrepreneurial urban agriculture in the United States. Lincoln Institute of Land Policy (2000)
17. Bertelsen, O.W., Zander, P.-O.: Obstacles to design in volunteer based organisations. In: Research Symposium, p. 93 (Year)
18. Nugent, R.: The impact of urban agriculture on the household and local economies. Growing Cities, Growing Food. Urban Agriculture on the Policy Agenda, 67–98 (2000)

Towards Engaged Consumption: New Sources of Inspiration for Eco-feedback Design

Stephen Snow and Margot Brereton

Queensland University of Technology, 2 George St, Brisbane, 4001, Australia
{steve.snow,m.brereton}@qut.edu.au

Abstract. Eco-feedback interventions are capable of producing reductions in household energy consumption. Yet less is known about exactly how this reduction is achieved, how to maximise user engagement, or how to effectively translate engagement into energy saving. This paper discusses design opportunities for eco-feedback systems through observations of domestic energy use in both Western and rural developing world contexts. Drawing on case studies from these two contexts including 21 empirical interviews, we present an alternative framework for human-resource interaction, highlighting design opportunities for a transition towards more engaged and sustainable energy consumption among users.

Keywords: Eco-feedback, resource use, electricity, engaged consumption.

1 Introduction

Domestic energy consumption continues to grow in saliency in the global consciousness due to its contribution to environmental issues such as climate change and the challenges associated with peak demand [10]. Providing households with improved feedback on their energy use (eco-feedback) has emerged as a fruitful means of addressing these challenges through heightening energy-awareness, better informing consumption decisions and potentially facilitating pro-environmental behaviours [3, 4, 5]. Research suggests that eco-feedback interventions such as energy monitors or interactive in-home displays are capable of producing significant reductions in household consumption, yielding average energy savings of approximately 7 to 14% for the time they are installed [3]. Less is known however, about the processes through which this conservation effect is achieved or which eco-feedback attributes best facilitate engagement and energy savings [6]. Similarly, while many eco-feedback studies assume a direct link between raised awareness and energy savings [6], social psychology literature suggests heightened awareness is not a reliable predictor of behaviour change [9]. What is clearer is that many Western consumers are substantially disconnected from the energy resources they consume [9] and that eco-feedback has the potential to help bridge this divide [5].

Adopting a broader frame of reference, one could justifiably question the significant amounts of time and effort being invested in this field by HCI researchers, when

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 117–124, 2013.

approximately one third of the world's population still rely on burning biomass as their primary source of energy [7]. Despite the peak demand issues being experienced in many parts of the world, 25% of the world's population still lacks effortless access to electricity [1] and cannot directly benefit from conventional eco-feedback. On the other hand, these populations unwittingly have a far stronger understanding and appreciation of the relationships between their everyday practices and their resource consumption. For instance, they physically gather the energy resources they require in many cases rather than simply flicking a switch or turning a tap.

Considering eco-feedback has the potential to better engage Western consumers with their energy consumption [5], we believe this field of enquiry has much to gain from studying practices of people in less developed countries who are already more engaged in this regard. This can serve to inform the design of eco-feedback technology and in turn, provide inspiration for designs that may benefit these communities as well.

The purpose of this paper is to investigate eco-feedback design through the lens of the human-resource interactions which take place in the home and the underlying infrastructures and mentalities that shape these interactions. Drawing from examples from both Western and rural developing world contexts, including 21 interviews with suburban Australian households, we illustrate the discrepancy in engagement with consumption between the two contexts. We then present a framework for human-resource interaction and highlight some of the many design challenges associated with eco-feedback effectively transitioning users towards more *Engaged Consumption*.

2 The Story of Two Worlds of Energy

2.1 The Western Context

The relationship between everyday practices and domestic resource consumption, whilst inextricable, is often poorly understood in Western contexts [10]. The continual disparity between the spheres of production and consumption represents a major obstacle to consumers becoming better engaged with their domestic resource use [9]. Electricity and water consumption for instance result from practices in the household sphere which can be unpredictable, changeable and based on social, cultural or external influences [10]. Yet in Australia, the billing mechanisms for these 'services' still consist of infrequent unitemised invoices which distill the multitude of actions creating the consumption into a single value. This has contributed to a mentality of disinterest among many suburban Australian consumers towards their resource consumption. Sofoulis [9] argues this is at least partly due to the longstanding dominant socio-technical systems for supplying water and electricity. In suburban Australia, large centralised utilities have assumed all responsibility for the supply and (in the case of waste water and sewage) disposal of these resources as a 'service' to their customers [9]. As a result of this centralised municipal system, an illusion is created of 'endless supply' at the other end of a power socket or tap [9]; a misconception that was actively promoted by electricity authorities in previous decades:

> *"Now, under the Council's new Tariff of Penny Units… You can use Electricity for every purpose without counting the cost... No need to worry about switching off the lights, but have peace of mind and contentment with a well-lighted home. – You can, without being extravagant, use electricity for everything!"* **(Electricity Topics, Issue 1, pp.8, 1936; reproduced in [8])**

The notion of endless supply is further reinforced by the typical home infrastructure related to electricity and water consumption. For instance the positioning of the water and electricity meters external to the dwelling emphasises that responsibility for the management of these resources lies outside of the home [9]. Many years of cheap and stable electricity prices have further reinforced this disengagement among suburban Australians with their energy consumption, thereby limiting motivation to better understand the factors affecting their consumption or adopt conservation practices.

Providing consumers with better feedback on their energy use represents a means of bridging this gap in energy-literacy and facilitating a more *Engaged Consumption* among consumers. However, considering the weak correlation between awareness and behavior change [9], a key feature of *Engaged Consumption* and thus a challenge for designers, should be to effectively translate engagement into energy conservation.

2.2 The Rural Developing World Context

Despite facing far more significant challenges, disconnection from resource consumption and lack of conservation awareness is not a problem facing many of the hundreds of millions of people living in rural areas in the developing world. For those who remain without effortless access to electricity and water, resource consumption is a far more tangible, visual, physical and social activity [1]. The link between everyday practices and resource consumption is more explicit, for instance, the same user may gather the resources themselves prior to overseeing and regulating their consumption [7]. In this context, feedback on usage is visual and immediate and often requires a prompt response, such as stoking a fire or filling a lamp.

Bidwell et al. [1] report on the design and deployment of solar-powered charging stations for mobile phones in the district of Mankosi on the Eastern Cape of South Africa. For many here, electricity is used for the single purpose of charging phone batteries and the supply source (the sun), the distribution network (wires) and the electricity consumption (phones being charged) can all be visualised within a single line of sight. This visualisation contrasts starkly with the act of plugging a charger into a socket on the wall, which many Western consumers are familiar with. In Mankosi, electricity consumption is often a very explicit social process; locals gathered at the mobile charging stations to chat with the charging facilitator and other locals. Users became aware of the factors affecting consumption such as cloud or rain through such informal conversations [1].

Another disparity between domestic resource consumption in the developing world and the West is the amount paid for electricity; both in terms of financial capital and physical effort. Using the example of Sagar Island in West Bengal, India [2], men and boys from villages walk to makeshift battery charging businesses to rent car batteries. These batteries are carried home to provide power for lighting, entertainment and other uses during the evening. For these simple privileges and the social status they afford, villagers exert considerable physical effort and pay the equivalent of 10 times the standard electricity tariff in West Bengal [2]. A similar situation exists in Mankosi [1] where, in the absence of the solar charging stations, villagers walk to establishments with petrol generators and pay around R5.50 (~US$0.60) to charge their phone when they can afford to. This cost is similarly exorbitant when compared to that incurred by charging a phone battery from a grid-connected dwelling in South Africa (~R0.70 or ~US$0.08)[1]. In both these examples, users have a high tacit knowledge of the factors affecting battery life and ways to conserve energy [1, 2].

3 Interview Findings – The Australian Context

In order to explore engagement between Australian consumers and their electricity consumption and to seek design inspiration for eco-feedback systems, we began by interviewing 21 participants living in Brisbane, Australia. While our findings here are limited to suburban Brisbane, over time we aim to balance these with accounts from remote Australian communities including primarily Indigenous communities whose remoteness often dictates they operate more self-sufficiently.

All of our participants lived in dwellings reliant on mains electricity and reticulated water and sewage, with 16 of the 21 having opted to receive a small digital energy monitor as part of a government sustainability initiative. We were not involved in the design or deployment of this monitor and cannot comment on the circumstances in which it was installed in the homes. The monitor displayed aggregate household consumption instantaneously both in kilowatt hours and in dollars. The interview process included a 'tour' of the dwelling and included questions covering the topics of electricity usage behaviour, electricity feedback and the social sharing of information related to electricity.

3.1 (Dis)engagement with Energy Consumption

Continuous electricity supply was seen by participants as both a necessity and a right. Despite this, the connection between understandings of the impact of everyday practices upon electricity consumption was found to be somewhat tenuous for many participants. For instance "turning off the lights" was mentioned frequently as a conservation practice, both in relation to the participants themselves and to teaching children. However, considering most participants had replaced their incandescent bulbs with energy saving compact fluorescent varieties, turning off lights often

---

1 Indicative only; assuming phone charger draws 3.68W while charging; phone takes 4 hours to charge; at ~R47/kWh. Sources: Eksom Schedule of Standard Prices for Local Authority Supplies, South Africa `http://www.eskom.co.za/c/article/1090/schedule-of-standard-prices/` and Lawrence Berkley National Laboratory `http://standby.lbl.gov/summary-table.html`

represented a less effective conservation action than modifications to other comfort, cleanliness or entertainment practices. In eight other instances, while participants were aware of an appliance consuming a lot of power, such as the air conditioning or tumble dryer, its usage was considered non negotiable. Three participants had signed up to a direct debit scheme where an agreed amount was automatically paid to the electricity utility each month, thereby distancing them even further from their bills.

> *"I don't even get a bill anymore; it's all on line…. I don't even see the little chart with your energy consumption displayed last year versus this year... I used to but not anymore, it just gets automatically paid"* **(Interview 1)**

3.2 Experiences of Simplistic Eco-feedback

Despite 16 of the 21 participants having received an energy monitor at some point, only half of these were found to be operable at the time of the interview. While four of these cases can be explained by technical malfunction caused by the install of solar panels, the other cases appeared to be due to neglect or could not be explained. For the remaining eight working monitors, at the time of the interview half were no longer referred to regularly or were found obscured behind other items (Figure 1).

Fig. 1. Two functioning electricity monitors obscured from view

Furthermore, despite the prevalence of the energy monitors among the sample group, all but one participant still relied on the quarterly bill as their main source of feedback on their electricity consumption. Several participants also had trouble reconciling their electricity bills with expectations of what they thought their bills "should" be, resulting in a disbelief or distrust of the utility responsible for generating the bill. These findings correlate closely with Sofoulis's [9] arguments regarding the implicit delegation of responsibility for electricity and water resources to external bodies.

> *"Sometimes I think the electricity utilities are dudding you all the time… I think they're estimating a lot and I don't trust them…. I have very set pattern of living. And I can't see how my power bill would triple. TRIPLE over a few short years. It doesn't make sense to me"* **(Interview 10)**

Despite this, all but one of the participants who had received the energy monitor claimed they had drawn some benefit from it at some point. In most cases however, engagement with the device had not been long lasting, or they had experienced difficulty relating the information displayed on it to specific actions or practices.

> *When we first got it I looked and it and I thought it was fascinating, but I didn't understand it. Like it just says, y'know, 0.68 or something and I went "that's interesting", but I couldn't relate it to (anything).... I know obviously the lesser the figure the better it is, but that's all.* **(Interview 20)**

Nearly all of the participants stated they would like additional feedback. In relation to their "ideal" electricity feedback system, the only attribute that was consistently mentioned as being desirable across the whole sample group was the provision of appliance specific feedback. This finding correlates with those of other HCI researchers in the field [5]. The desirability of many other attributes, for instance spatial comparison, social comparison, positioning and means of accessing feedback varied, with these attributes being desirable to some participants and undesirable to others.

3.3 Conversations Related to Energy Consumption

Finally, we found an interesting disparity between participants' accounts of their conversations regarding (1) electrical appliances that contribute towards energy consumption and (2) absolute measures of consumption such as electricity bills. While appliances were discussed by most participants in social circles and between friends, especially prior to making a purchase; only six of the 21 interviewees mentioned they shared their bill amount with anyone outside of their immediate family. This does not mean that the bill was considered a secret or private thing, as only three participants described it as such, but simply that conversations regarding measurements of electricity consumption were generally contained within the family home, whereas conversations about appliances and their use were shared.

4 Towards Engaged Consumption – Design Opportunities

Through our case studies and interview findings, this paper illustrates a considerable disparity in engagement with domestic resource use between suburban Australian and rural developing world contexts. Therefore we raise the question of: what lessons can be learnt from resource use in the developing world toward better engaging Australian consumers with their consumption?

To address this, we highlight design opportunities for a transition towards a new framework for human-resource interaction involving more *Engaged Consumption* among consumers. Here we draw inspiration from our case studies, interview findings and findings made by other researchers. Table 1 presents an overview of our framework and we explain some important characteristics in better detail below.

Table 1. A framework for Human-Resource Interaction comparing current experience in Mankosi, West Bengal and Suburban Australia with the vision of a more informed and environmentally aware *Engaged Consumption*

Human/Resource Interaction	**Mankosi / Sagar Island**	**Suburban Brisbane**	**Engaged Consumption**
Effort expended to access energy resource	High: Walking; gathering; carrying	Low: Flicking a switch, turning a tap etc.	Low/Medium: Effort required to purchase electricity credit and to monitor energy use through feedback
Perceived Benefit/ value	Luxury item; social status and connection; more usable hours in the day	Assumed necessity; Maintenance of lifestyle	Informed and reduced usage, cost; better control over consumption
Awareness of resource consumption, cost	High: Resource consumption is tangible; cash paid for given quantity of electricity prior to usage	Low: Little appreciation of resource use; infrequent bill; no point of sale information	Higher: Appliance and practice-specific feedback on usage; cumulative bill
Sociability of resource consumption	High; talking while charging, walking; sharing information	Low; bills discussed mainly within the home	Higher: more social learning about electricity

Based on Strengers [10], we consider a key attribute of *Engaged Consumption* is eco-feedback that enables users to relate their resource consumption to everyday practice. This is something that is more easily done at present in Mankosi or Sagar Island than in suburban Australia. For instance, in developing world contexts, paraffin may be used for lighting; wood or coal used for heating and cooking; and electricity for entertainment and communication [7]. In many of our participants' dwellings however, electricity was used for all of these purposes. As such, one possible design strategy for eco-feedback systems could be to provide people with both appliance-specific (e.g. microwave, cooktop) and practice-specific (e.g. cooking) feedback on usage.

In relation to *Engaged Consumption* effectively facilitating energy conservation, we highlight three aspects of resource use in the developing world which serve to shape conservation. These include the physical effort required to obtain resources, the timing of payment for them and the consequences of wastage. On Sagar Island for example, electricity use is defined by the physical effort and high cost involved with the battery rental process [2]. Money is paid for a given quantity of electricity prior to its use and a direct consequence of injudicious usage is no electricity for the remainder of the evening or further physical effort and money spent renting a new battery. For many of our participants however, there were no clear indicators of electricity wastage, no physical consequences of wastage and the economic consequences would not become apparent to them until the next quarterly bill.

Considering it may be unrealistic to introduce physical activity as a pre-requisite for access to electricity in Western contexts, we focus on two more achievable reforms to human-resource interaction aimed at translating engagement into energy conservation. These are: (1) better feedback on inefficient use and; (2) bringing the physical payment for the energy closer to the point of consumption. We consider introducing or at least simulating (through eco-feedback) a system of pre-payment for energy resources to be a key component of *Engaged Consumption*. In a pre-payment scenario, the user 'gathers' credit and has an obligation to monitor and control usage such that their credit, along with their power or water supply, does not run out before the credit is recharged. Eco-feedback is fundamental in any pre-payment scenario by providing users with information on their consumption and their remaining credit, thereby positioning price at the forefront of consciousness and counteracting the notion of 'endless supply'. Opportunities exist here for eco-feedback that delivers powerful visual representations of money dripping away or energy credits disappearing. Furthermore, utility-led systems for energy pre-payment are already available in areas of both the Western and developing world and when combined with eco-feedback have produced impressive energy saving results [3].

A final design opportunity that we highlight as a potential means of achieving *Engaged Consumption* is that of sharing information related to electricity. While electricity use was found to be a highly social process in Mankosi [1], we found that less than one third of our interviewee respondents shared any billing or usage information with those outside of their immediate family. We conclude by encouraging HCI designers to look for opportunities for how eco-feedback can better facilitate information sharing and social learning about energy use between friends and family.

References

1. Bidwell, N.J., Siya, M., Marsden, G., Tucker, W.D., Tshemese, N., Gavin, S., et al.: Walking and the social life of solar charging in rural Africa. In: Proc: TOCHI (2013)
2. Dadiomov, D.: Power trips. Stanford Magazine (May/June 2012), `http://alumni.stanford.edu/get/page/magazine/article/?article_id=53335` (accessed March 20, 2013)
3. Faruqui, A., Sergici, S., Sharif, A.: The impact of informational feedback on energy consumption-A survey of the experimental evidence. Energy 35(4), 1598–1608 (2013)
4. Froehlich, J., Findlater, L., Landay, J.: The design of eco-feedback technology. In: Proc: CHI 2010, Atlanta, Georgia (2010)
5. Froehlich, J., Findlater, L., Ostergren, M., Ramanathan, S., Peterson, J., Wragg, I., Larson, Fu, F., Bai, M., Patel, S.H., Landay J.: The design and evaluation of prototype eco-feedback displays for fixture-level water use data. In: Proc: CHI 2012, Austin, Texas (2012)
6. Hargreaves, T., Nye, M., Burgess, J.: Making energy visible: A qualitative field study of how householders interact with feedback from smart energy monitors. Energy Policy 38(10), 6111–6119 (2010)
7. Howels, M.I., Alfstad, T., Victor, D.G., Goldstein, G.: A model of household energy services in a low-income rural African village. Energy Policy 33, 1833–1851 (2005)
8. King, J.: Power for the people: Brisbane electrical heritage 1880s-1950s. Published by Queensland Energy Museum, Brisbane (2003)
9. Sofoulis, Z.: Big water, everday water: A sociotechnical perspectice. Journal of Media and Cultural Studies 19(4), 445–463 (2005)
10. Strengers, Y.: Bridging the divide between resource management and everyday life: Smart metering, comfort and cleanliness. RMIT University (2009), `http://researchbank.rmit.edu.au/eserv/rmit:6820/Strengers.pdf` (accessed March 19, 2013)

Sustainability at Home: An Exploratory Study on Monitoring Needs and Energy Management Actions of Solar Power Producers

Dounia Lahoual[1] and Myriam Fréjus[2]

[1] C3U/Paragraphe Laboratory - Université Paris 8, 2 Rue de la Liberté, Saint-Denis, France
dounia.lahoual@gmail.com
[2] EDF Research & Development, 1 Avenue Général de Gaulle, Clamart, France
myriam.frejus@edf.fr

Abstract. This exploratory study focused on the energy consumption practices of customer-producers (prosumers) in relation to their needs in monitoring energy production. Our analysis of both production monitoring activities and domestic activities in real situations revealed the motivations of these producers and demonstrated that the actions of energy management were not dependent on the status of customer-producer. The actions of energy management arose from individual and collective constructions, as well as the appropriation of electrical appliances and attractive pricing offers. These results suggest that the issue of offering incentives for energy management would benefit from greater attention to questions of appropriation, pricing, and technical devices.

Keywords: Sustainability, energy management, photovoltaic production, domestic activity, prosumers, consumption/production feedback.

1 Introduction

Climate change and the growing awareness of our limited resources have made energy management a key issue in today's world. For energy utilities, reducing peak demand and incorporating renewable energy sources (wind turbines, photovoltaic panels, etc.) have become priorities. Electrical systems should be operated more flexibly, and constraints due to intermittent energy production, unpredicted production and the decentralization of green energies need to be managed. To address these issues, transmission and distribution networks have favored the creation of smart grids which depend heavily on the deployment of information and communication services and technologies [3, 7]. The European Grid4EU project, which includes the French NiceGrid project, was launched in 2012 in this context. The NiceGrid team is building a demonstrator of a smart solar district that incorporates both high photovoltaic energy production and storage batteries to allow each household to store energy during peak production. This project also enables the customer to balance their energy use according to their production, while taking into account the overall load on the grid. The goal of the energy utilities is to place the

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 125–132, 2013.

customer in the role of actor. The utilities suppose customers will become “prosumers” who actively manage their energy consumption [14, 15]. For these reasons, utility companies wish to carefully follow them (monitoring software, assistance in system installation, understanding of appropriation processes) in order to anticipate the impact of their energy production on the grids.

1.1 The Study of Production Monitoring Activities and Consumption Practices of Energy Producers

Underlying this project is the notion of prosumer practices and their potential impact, a field of research having much in common with the more widely investigated field of behavior modification in human-computer interactions, particularly with regard to the design and evaluation of "consumption feedback" [4, 8, 9]. Several studies provide evidence for the hypothesis that energy-producing households are more inclined to actively manage energy consumption [17, 20]. Various means to influence consumption practices have been specifically designed, taking into account both the load on the grid and the anticipated production [13]. Moreover, studies of household energy producers suggest that they employ supervisory [10] and maintenance [6] routines. Although many studies have focused on the social acceptance of renewable energy and the emergence of local energy communities [1, 19], few have investigated prosumers, their motivations to become producers, and how they purchase solar panels. The activities of energy consumption, production monitoring, and the possible links between the two have not been sufficiently explored.

1.2 Situating the Activities of Prosumer Energy Management in the Domestic Context

Studies [20] have emphasized that involving householders in environmentally friendly initiatives requires not only changing intrinsic motivations, but also making much broader systemic interventions that address the circumstances surrounding the desired activity. It is well known that energy consumption practices cannot be influenced merely through the provision of information and incentives, because these practices grow out of domestic activities devoted to other objectives, such as home maintenance, meal preparation, entertainment, and so on [16]. It is therefore important to situate energy consumption in this context and to analyze the determinants of individual and collective home activities.

Our exploratory study sheds light on these issues by focusing on the following:

- The specific needs of energy producers: What information and what types of interactions are most useful to them? Do these producers carry out specific activities in managing their production? Do they have specific needs for assistance in making decisions, especially through interactive and/or informational devices?
- The relationship between energy production and consumption: Does this new status of customer-producer generate specific activities related to energy management

or to the balance of production and consumption? Does production monitoring influence energy management?

2 Methodology

2.1 Data Collection

Study Population. Six photovoltaic (PV) energy producers were recruited for this exploratory study (Table 1). The homes were equipped with solar panels and electric heating, and the households were at different stages in appropriating the production management system, which allowed us to carry out a diachronic analysis.

Table 1. Characteristics of the participants and summary of collected data

Age of participants	Producers since	Profession of solar power producers	Family composition	Exploratory interviews	Collective self-confrontation interviews	Video recording
① Mr L. (45 years old)	2 years	Executive officer in tourism	Married with one young child (3 people)	1h23	2h44	8h
② Mrs N. (42 years old)	6 months	Household electrical saleswoman	Divorced with two teenage daughters (3 people)	1h53	1h05	8h
③ Mr F. (57 years old)	1 year & 5 months	Retired (former train driver)	Couple (2 people)	2h43	2h14	8h
④ Mr P. (61 years old)	1 year & 7 months	Retired (former computer engineer)	Couple with two teenage children (4 people)	1h22	/	/
⑤ Mr Ca. (50 years old)	2 years & 6 months	Teacher	Couple with three children (5 people)	38min	/	/
⑥ Mr Co. (52 years old)	3 years	Inactive (previously wine-grower)	Couple with one child (3 people)	15min	/	/

Interviews. We first carried out exploratory telephone interviews with three of the producers (4, 5, and 6 in Table 1) in order to determine their motivations to become energy producers, their personal projects, their patterns of energy consumption, and their energy management actions. These elements aimed to complete the bibliographical research in order to refine the methodology of the study and the recruitment criteria of participants.

Collecting the Audiovisual and Written Data on Domestic Activities. To reconstruct the situated dynamics of domestic activity, our study was conducted within the theoretical and methodological framework of the Course-of-Action research program. The Course-of-Action methodology allows grasping the meaning each householder gives to his or her own activity [11, 18]. We filmed three of the households (1, 2, and 3 in Table 1) in order to gain insight into their individual and collective activities (the long duration of recruitment limited us to three households). The recording set-up was as follows: unobtrusive mini-cameras with a 120° view angle and equipped with microphones were programmed to begin recording in our absence. Several switches allowed the households to interrupt the sound and video recording if needed. The time and the place for recording were scheduled after discussion and validation by the households (formalized by the signature of an agreement). Several rooms with high activity levels were chosen for recording: the kitchen, the living room and a main hallway where the residents mainly circulated. For each room, we scheduled continuous recordings of two hours in four moments in the week in order to collect various domestic activities (with the household's agreement). In addition, every day the residents used a notebook developed for this study to provide requested information on the use of a range of electrical appliances. Both notebooks and video recordings were used in the self-confrontation interviews in order to grasp the meaning they attributed to their activities in situation (viewpoints, motivations and goals). To achieve this, we selected recording and notebook sections on the basis of clearly defined criteria: the chronological aspect and the unexpected nature of domestic activities, the presence of energy management actions and activities mediated by electrical devices. The households were compensated for participating with discount vouchers.

2.2 Data Analysis

Data Transcription. The video recordings of the home activities were transcribed using the multi-score method [5]. This method allowed us to summarize the inhabitants' actions from their individual viewpoints and the chronological use of electrical appliances in relation to the space they occupied (see Table 2).

Constructing Shortened Stories. We constructed shortened stories, which were chronological presentations of the meaningful units of activity in which the inhabitants were engaged [12]. These narratives were constructed from the multi-score transcriptions.

Analysis. The analysis sought to model production monitoring activities and the instantiation of control activities in the domestic courses of action. Concerns and the resulting generation of action from each individual's point of view were described and compared, as were items of knowledge mobilized for the action and any meaningful elements that were used to act.

3 Results

3.1 Identifying Activities Linked to Monitoring PV Energy Production

Motivations of the Customer-Producers. The PV producers had two main motivations. The first was financial and, in this sense, the financial aid given to install the panels and the guarantee that all generated energy would be purchased by the supplier were a determinant. Energy self-sufficiency over the long term was also an important consideration with regard to an eventual end to the supplier's guaranteed buy-back price and the progressive rise in electricity costs.

Producers Specific Activities. After the system was installed, several activities related to production monitoring were noted. All the producers verified that the system was functioning correctly, that enough energy was being produced, and that technical defects or breakdowns had indeed not occurred. These activities were carried out with the support of several devices: an inverter (with light or sound displays representing the functioning of the inverter) used by three households (1, 2, and 3 in Table 1), a production meter (kWh on the display) used by only one producer (6), and monitoring software (the default setting being the daily production graphics) used by two producers (4 and 5). In addition to ensuring the proper functioning of the PV system, three participants (3, 4 and 5) performed another activity, though this was less frequent: tracking the energetic and financial profitability of the system. In the first case, two participants (3 and 4) with varying frequency, monitored the system efficiency by taking readings from the inverter in kWh of production and transferring this data to Excel spreadsheets for personalized and ongoing tracking (e.g., converting to Euros, comparing energy produced to consumed, etc.). Tracking the financial profitability (done by producer 3) involved the same activities but the goal was slightly different: trying to determine the point in time at which the financial gains from selling energy would be greater than the financial outlay for the installation, thus ensuring a regular income. In addition to these monitoring activities, participation in online discussion forums was a veritable resource for support and problem-solving during the installation phase. Three producers were involved in forums (4, 5 and 6). Two distinct profiles of these energy producers nevertheless emerged from our analysis: three households (1, 2 and 6) carried out light, routine monitoring mostly to simply ensure proper technical functioning and the other half (3, 4 and 5) who were much more involved. These latter monitored the system much more closely and systematically and clearly sought profitability and/or energy self-sufficiency. They were also much more invested in participating in the specialized internet forums.

3.2 Domestic Activities and Energy Management

Actions of Energy Management. The analysis of the domestic activities revealed a repertory of actions for energy control. These actions arose sporadically in the midst of the domestic activity and usually indicated a concern about energy use during the course of the activity (reducing energy costs, avoiding waste, taking advantage of an energy resource already in use, etc.). Several forms of energy control were identified in the data: shifts in the times of use, grouping energy uses together, finding shortcuts, voluntarily avoiding energy use, turning appliances off, insulating, and making targeted purchases (certain of which were also noted by Pierce and al. [9]). These actions were fully incorporated into the domestic activities, such as systematically turning off an electrical burner before the food was completely heated in order to take advantage of inertia (household 1 in Table 1), turning lights off on leaving a room (all the participants), letting natural light in rather than turning on electrical lights (1, 2 and 3 in Table 1), etc. These gestures were both automatic and easy to do. They did not guide any of the domestic activities but rather were performed sporadically. These control actions were also built up over time. In some cases, incentives prompted these actions. For example, the option and lower cost of running appliances at off-peak times encouraged a shift in the time of using washing machines and dishwashers (1, 3, 4 and 5). In other cases, personal experience prompted change, as for the homemaker who had lived in Cuba (1), where energy constraints are high. She had noted a sharp rise in her electrical bill when she ironed clothes and thus had developed the habit of grouping together her ironing into single sessions and then spacing the ironing sessions out. This person displayed a culture built up around the links between energy usage and energy cost and she thus also grouped stovetop cooking around a single burner: she had determined that turning off an electric burner and then turning it on again was more costly. These control actions were shared by all household members and could in some cases be described as co-realized. For example, the couple (1) coordinated doing the laundry, with the woman preparing a laundry load during the day and the man starting the washing machine in the evening during the off-peak period. All in all, we observed a plurality of uses, including three (shifting, grouping and shortening) which illustrate the appropriation of electrical devices. The analysis thus indicated that, to encourage home energy management, household members need to fully appropriate their electrical appliances, which means that they are free to use them as they see fit with regard to their home activities, external parameters (weather, climate), and the comfort level they are seeking. Indeed, the observed control actions revealed that households do not necessarily use the electrical appliances as initially envisioned by the designers. It also implies that they have access to the services that support their actions (pricing, management systems). The actions of energy management should not interfere with any of their normal home activities.

Energy Producers: Customers as Actors in Their Own Consumption? The quantity of generated solar power is not a criterion to guide consumption because, first of all, in the households we observed consumption was on average about ten times greater than the PV production. Moreover, we observed that the actions of energy management resulted from the appropriation of electrical appliances and attractive price offers, which were independent of the status of producer. These actions reflected individual and collective constructions of appropriation and could surely be carried out by all customers, given favorable conditions and the means to integrate them into more general domestic practices. We thus confirm that the actions of energy management are not linked to the status of customer-producer: you become a producer because you are already an actor – that is, because you are already involved in energy management – and not the reverse. The shift to producer is a step in an overall project of energy control, where the development of actions to manage energy consumption and the ability to use available resources are prerequisites.

4 Conclusion and Perspectives

This exploratory study provides two orientations for the design and development of future research: the design of systems to visualize and monitor energy consumption in relation to production and the further investigation of how the actions of energy management develop. Within the framework of monitoring design, this study indicates the types of information that should be displayed by inverters, the functions that are most useful in monitoring software, and the types of energy information that should be available inside the home. Specifically, we have determined that monitoring tools should target long time scales linked to the expectation of profitability and daily and monthly analyses. The monitoring system should not offer only information on real-time data. In fact, to encourage energy management, it is crucial to facilitate householders' appropriation of technical devices, electrical appliances and advantageous pricing offers that orient the organization of domestic courses of actions. Domestic activities could be assisted by automatic systems that take over for the residents (starting the washing machine during off-peak hours). However, any actions of automatic energy management need to be transparent to the household members and should not cause any disruption of household routines. The question of how to coordinate automatic control systems with ongoing home activities (distribution of automatic operations, visibility and easy recourse to manual control) is thus clearly posed, pointing back to a fundamental issue of human-computer interaction design. Last, this study suggests that the debate on energy management will be enriched by a shift of emphasis toward questions of appropriation, which we believe are essential [2]: we need to think about designing future situations of energy management in terms of those situations that are appropriable. Other studies are needed for us to instantiate this objective in the design and evaluation of the neighborhood solar project we are currently engaged in. Our future studies will be longitudinal studies in order to understand how individual and collective behaviors and energy management practices are built before and after the insertion of production systems.

References

1. Alvial-Palavicino, C., Garrido-Echeverría, N., Jiménez-Estévez, G., Reyes, L., Palma-Behnke, R.: A methodology for community engagement in the introduction of renewable based smart microgrid. In: 2011 International Energy Initiative. Elsevier Inc. (2011)
2. Dix, A.: Designing for appropriation. In: Proc. HCI 2007, pp. 27–30 (2007)

3. Fan, Z., Kalogridis, G., Efthymiou, C., Sooriyabandara, M., Serizawa, M., McGeehan, J.: The New Frontier of Communications Research: Smart Grid and Smart Metering. In: Proceedings of the 1st International Conference on Energy-Efficient Computing and Networking (2010)
4. Froehlich, J., Findlater, L., Landay, J.: The design of eco-feedback technology. In: Proc. CHI 2010 (2010)
5. Guibourdenche, J., Vacherand-Revel, J., Grosjean, M., Fréjus, M., Haradji, Y.: Transcribing the activity of a distributed family. In: Proc. CSCW, pp. 637–640. ACM DL, China (2011)
6. Juntunen, J.: Domestication of small-scale renewable energy systems – A case study of air heat pumps, residential micro wind stations and solar thermal collectors in Finland. In: EcoDesign 2011 (2011)
7. Lampropoulos, I., Vanalme, G., Kling, W.: A methodology for modeling the behavior of electricity prosumers within the smart grid. In: Proc. of the 2010 IEEE PES Conference ISGT Europe (2010)
8. Pierce, J., Fan, C., Lomas, D., Marcu, G., Paulos: Some consideration on the (in)effectiveness of residential energy feedback systems. In: DIS 2010 (2010)
9. Pierce, J., Schiano, D., Paulos, E.: Home, Habits, and Energy: Examining Domestic Interactions and Energy Consumption. In: Proc. CHI 2010 (2010)
10. Pierce, J., Paulos, E.: The Local Energy Indicator: Designing for Wind and Solar Energy Systems in the Home. In: Proc. DIS 2012 (2012)
11. Pinsky, L., Pavard, B.: What Does Real Work Analysis Tell Us about System Design? In: van der Veer, G.C., Green, T.R.G., Tauber, M.J., Gorny, P. (eds.) Readings on Cognitive Ergonomics, Mind and Computers. LNCS, vol. 178, pp. 248–259. Springer, Heidelberg (1984)
12. Poizat, G., Fréjus, M., Haradji, Y.: Analysis of collective activity in domestic settings for the design ubiquitous technologies. In: Proc. ECCE. ACM DL (2009)
13. Schrammel, J., Gerdenitsch, C., Weiss, A., Kluckner, P.M., Tscheligi, M.: FORE-Watch – The clock that tells you when to use: Persuading users to align their energy consumption with green power availability. In: Keyson, D.V., et al. (eds.) AmI 2011. LNCS, vol. 7040, pp. 157–166. Springer, Heidelberg (2011)
14. Shandurkova, I., Bremdal, B., Bacher, R., Ottesen, S., Nilsen, A.: A prosumer oriented energy market: developments and future outlooks for smart grid energy markets. Improsume Publications Series (2012)
15. Shuen, A.: Web 2.0 A Strategy Guide (2008), `http://www.scribd.com/doc/15490958/Web-20-A-Strategy-Guide-by-OReilly-Media` (retrieved June 14, 2010)
16. Stern, P.C.: What Psychology knows about energy conservation. American Psychologist 47(10), 1224–1232 (1992)
17. Strengers, Y.: Designing eco-feedback systems for everyday life. In: Proc. CHI 2011 (2011)
18. Theureau, J.: Course of Action Analysis & Course of Action Centered Design. In: Hollnagel, E. (ed.) Handbook of Cognitive Task Design, pp. 55–81. Lawrence Erlbaum Ass., Mahwah (2003)
19. Walker, G., Devine-Wright, P.: Community renewable energy: What should it mean? Energy Policy 36, 497–500 (2008)
20. Woodruff, A., Hasbrouck, J., Augustin, S.: A Bright Green Perspective on Sustainable Choices. In: Proc. CHI 2008, Florence, Italy, pp. 313–322 (April 2008)

WattsBurning on My Mailbox: A Tangible Art Inspired Eco-feedback Visualization for Sharing Energy Consumption

Filipe Quintal, Mary Barreto, Nuno Nunes, Valentina Nisi, and Lucas Pereira

Madeira Interactive Technologies Institute, University of Madeira, Caminho da Penteada, Funchal, 9020-105, Madeira, Portugal
{filipe.quintal,mary.barreto}@m-iti.org, {njn,lucas.pereira}@uma.pt, valentina.nisi@gmail.com

Abstract. This paper describes a novel art-inspired tangible eco-feedback system. The concept emerged from a workshop with researchers, designers and artists looking at innovative ways to provide more effective eco-feedback that engages users emotionally. The tangible aspect of the system is composed of a set of magnets that users can stick on their physical mailbox outside of their apartment building according to their average energy consumption. The magnets are a total of seven pieces, one for each day of the week. Each piece has a variation of three colors, from green (low consumption) to burning red (high consumption). The magnets are to be displayed in a sequence that represents a typical panorama of local nature. In this paper we report the design and the study we conducted to gauge preliminary results on the system usage and potential. Interviews with participants revealed that none of them felt uncomfortable having their consumption displayed outside. When children were involved in the process they "took control" of the task and pressured their families to perform better.

Keywords: Sustainability, Aesthetics, Art driven Eco-feedback, User Interfaces, Prototyping.

1 Introduction

Individual household consumption accounts for a significant part of the total worldwide consumption. For example, two thirds of the electricity used in the United States, 36% of the greenhouse gasses, and 12% of the fresh water consumption [1]. Most people are in fact concerned about the consequences of their actions, however they are also unaware of the impact of their daily activities and more importantly how they can change their behavior to reduce resource consumption. This gap between peoples concerns and behaviors, led to the creation of eco-feedback technologies. Eco-feedback is defined as "technology that provides feedback on individual or group behaviors with a goal of reducing environmental impact"[2]. A challenge when designing feedback about energy consumption, is to relate current actions with the

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 133–140, 2013.

impact that they can have in the world, commercial energy feedback systems normally present that information as g of CO2.

Pierce et al [3] surveyed several publications exploring the impact of eco-feedback technologies in energy consumption and on consumers' behaviors. The authors found out that the use of eco-feedback technology resulted on savings between 5 and 12% of daily energy consumption. They also concluded that when savings didn't happen the eco-feedback was displayed too infrequently (monthly) and hence was disconnected from the consumption behavior. Furthermore studies have show that the savings can relapse over time as the consumers' interest in the feedback device decrease [4]. Another important finding is that the normal representation of kWh is not informative enough to the majority of the users. Furthermore in [5] the authors argue that this lack of understanding made users remove those metrics from proposed eco-feedback designs.

In fact the entire electricity infrastructure is designed in way that is hidden to most consumers. With WattsBurning on my mailbox, we want to make the household energy consumption somehow tangible to the users and foster a sense of awareness about it through the action of updating it every day and sharing with their neighbors. By adding it to users routines, we plan to explore the accountability felt by the users.

In the remaining sections of the paper we will present the related work and the WattsBurning concept. In the last sections we describe the deployment of the study, its analysis and the following discussion and conclusions.

2 Related Work

2.1 Art Inspired and Tangible Eco Feedback

Pierce and Paulos [6] argue that feedback systems are to be designed as some*thing tangible* as a starting rather than ending point for design inquiry and exploration. That raises cognitive awareness and motivates energy conservation behaviors. In order to provide for this lacking experience, the authors developed Energy Mementos, treating energy as a material that is experienced in a unique and meaningful way. The STATIC! Research institute has developed several tangible eco-feedback devices. Energy Curtain attempts to motivate people to use more sunlight by opening the curtains [7]. The feedback is given in colored patterns in a window shade woven. PowerCord, a power aware cord, which was designed to visualize the energy consumed opposed to hiding it. Power Cord glows according to the level of energy that is passing trough [7]. Johnson et all [8] developed Watt-Lite a tangible display of electricity consumption for the workplace. In Watt-Lite, consumption is displayed by the projection of 3 torches to the floor, one represents the real time consumption, while the other 2 represent the maximum and minimum of the day. Using a more abstract approach artists Lena Bllimeier and David Baur, designed the Vision Energy, which is a set of sculptures in order to to "transform the abstract energy consumption process in to a visual adventure" [9]. In [10] digital artist Tiffany Holmes reports on her project projecting the visualization of the energy consumption of households in the public space of a building hall (7000 oaks and counting). In Holmes case, digital art is used to display hidden data of real time usage of key resources such as electric appliances in order to offer new strategies to visualize energy in the home and workplace. Another relevant work is presented in [11] an art installation in Helsinki, in which the city energy consumption was displayed with lasers projected to the smoke of a coal energy power plant. During 7 days the smoke cloud was projected with green colors (the projection was actually the outline of a cloud), the lower the consumption the larger the illumination. In the final day residents were asked to unplug the devices to reduce consumption, to increase the size of the green cloud. This resulted in a reduction of the peak demand in 800 kVA.

2.2 Socially Enabled Eco Feedback

Researchers have used motivators such as social praise or peer pressure to motivate in consumption savings [12]. In [13] the authors designed a system to work with the Watson energy monitor[1]. During the study a group of participants were able to consult others consumption while other participants could only view their own. The results showed that the second group managed bigger savings when compared to the fist one, the research team observed that participants were motivated to save energy by the element of competition. A similar study was performed with a mobile energy-monitoring tool, the research team observed that participants that shared their energy consumption with others, exchange expertise and troubleshoot problems between themselves [14]. Users engage in comparing and analyzing the others consumption mostly by curiosity.

Odom and Pierce [15] describe the design and implementation of an eco-visualization design to be used in the Energy Challenge, which was a competition in Indiana University in which students were competing to save water and electricity. The competition resulted in an estimated combined savings of 33.008 KWh of electricity and 724,322 gallons of water compared to baseline consumption. Additionally the research team observed that the consumption behaviors were strongly influenced by their peers, and that social motivation was the key component to the success of the competition.

3 System Description

3.1 Household Eco-feedback System

The study present here is part of a larger eco-feedback project, which provided families with consumption information via an android tablet. This research platform implements a low-cost non-intrusive sensing infrastructure that is capable of measuring energy consumption and detecting detailed appliance-level data. The consumption information is then made available in the cloud. It allows users to check their real-time or historical consumption data anywhere as long as an Internet connection is available. The system operates in two different modes (see Figure 1), the first one

[1] `http://www.diykyoto.com/uk/`

displays real-time data about the consumption (in kWh, €, CO_2 emissions), and the second mode is triggered when it goes to idle mode (normally after there is no interaction with device for more than 2 minutes). The second mode comprises a set of digitally manipulated pictures of a local forest landscape that occupies all the screen of the tablet. The landscape picture can change from very healthy green with rainbows and animals, to very sickening dark with fire and without animals, according to how the current consumption compares with the average of a homologous period.

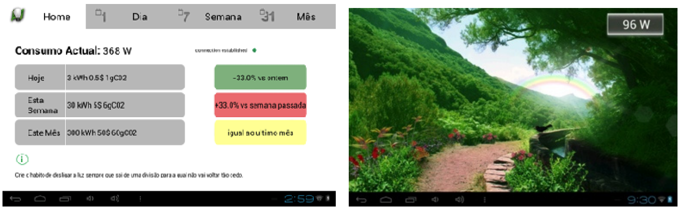


Fig. 1. Left:First mode of the tablet application; Right: and second operation mode.

4 WattsBurning on My Mailbox

The concept used in WattsBurning on my mailbox emerged from a workshop where researchers, designers and artists came together to think about innovative ways to provide eco-feedback beyond the traditional informative kWh or CO2 emission displays. Some of the workshops members had previous experience with public communal eco-feedback, in particular the TidyStreet initiative, which was part of the Change project [16].

The brainstorming session revolved around the combination of public (communal) display and using forms of artistic display connecting the energy consumption to elements of the environmental heritage of the local place. Several ideas emerged from the workshop most of the ideas looked at tangible ways of providing and sharing the eco-feedback from the households. The research team agreed that art and playfulness could play a major role in increasing the emotional connection between people, their energy consumption and the long-term effect on the environment. Starting from these ideas the team looked at public spaces that could be used to share the consumption eco-feedback. Since the system (described in 3.1) was deployed on residential apartment buildings the physical mailboxes provided an interesting place to share the consumption, which was both accessible and visible to the larger community. The mailboxes could act as a placeholder for a tangible display of energy information that was at the same time shared with their neighbors. In order to leverage the connection between the in-house system that was already installed and the new shared eco-feedback, the research team brainstormed on different ways to bring the artistic visualization of the consumption. Ideas ranged from using a portrait of the family with different moods, or use a local newspaper to display the consumption, and were finalized on a specific concept: mapping a forest landscape to physical magnets. For the mailbox magnets we used the same forest landscape that was used in the tablet eco-feedback application. (see Fig. 1)..The final concept was taken into different prototyping and evaluation cycles before emerging as the final WattsBurning on my mailbox concept.

4.1 Mailboxes Magnets

The system provides an opportunity for family members to share their energy consumption information they already accessed through the android tablet at home, using a shared physical space outside of their apartments. Both eco-feedback visualizations mapped household consumption to the natural endemic forest, which we knew from previous research provided a strong emotional connection in particular with locals and due to several recent extreme environmental phenomena like heavy rain and mudslides or severe forest fires [17].

The hardware used to implement the outdoor visualization was based on a set of magnets that participants placed in their mailbox according to their daily consumption. Each magnet represents the consumption of a specific day of the week, referring to part of the full illustration of the local forest at three different health stages: “sick”, “normal” and “healthy”. The landscape depicted here is a horizontal portion of the landscape present in the tablet system which users were already familiar with (See Figure 1 right). In total each family was given 21 magnets, three for each day of the week with the three different health levels. In addition there was also a banner placed above the magnets with the caption for each day (see Figure 3). Since the mailboxes were made of aluminum (the magnets did not stick to it) a small-magnetized rubber ban was placed in the mailbox through hooks so that the magnets could “stick” to it. Figure 2 shows the banner and the three different landscapes that created the set of magnets for the three different consumption levels. Figure 3 shows how the magnets were placed in the mailboxes.

Fig. 2. The 3 different stages of the forest represented by the magnets (banner is on top). 1. Banner with the week days. 2-Forest with low consumption. 3-Forest with medium consumption 4- Forest with high consumption.

5 Evaluation

To evaluate how sharing energy with the mailbox system impacted the perception and energy consumption we selected 4 families for a period of one week to test the concept. The sample families for this study were a sub sample from the families selected for the bigger deployment of the tablet based eco-feedback system (section 3.1). Therefore the families were already used to get information about their energy consumption, and they were aware of how their routines impacted the depletion. All the families were couples with children; in 3 of them the kids had less than 12 years. The families lived in two apartment blocks from the same building complex in an urban area of a small city from a southern part of Europe. The mailboxes were placed in the entrance of each apartment block and visible by passers by and neighbors. Users were informed about this study by phone and then scheduled for an interview later to explain how the mailbox system could be used.

5.1 Methodology

The mailbox study lasted one week, which we knew, from previous research, was enough to span the routines of a family impacting their energy consumption. Each morning at 7:00 AM the participants received a text message with information about their consumption in the previous day. The text message followed the next template: *Good morning Mr/ "Participant name" your consumption yesterday was above/equal/bellow your average for that day of the week, you should proceed to place the magnet assigned for high/average/low consumption.*

Then we expected family members to proceeded and place themselves the magnet in the corresponding position at the mailbox outside of the apartment building. At the end of the afternoon one of the researchers passed by the building to check if the participants had placed the magnets correctly (the participants were not aware of this verification). In the following week, we visited the families and conducted a short interview. The goal of the interview was to qualitatively verify if the metaphor used to represent the consumption was well understood, how the family performed the task, and how families felt about having their consumption displayed outside their houses.

5.2 Analysis

From our afternoon observations we confirmed that only one of the families skipped the magnet placement in one day. However they placed two magnets in the following day. Apart from that case all the families performed the task on time and correctly (the placed magnet was the correct according to the consumption).

All the families found the task of getting the message and placing the magnet simple and reported it didn't interfere with their normal daily routines. They also found the magnets easy to understand. None of the families felt discomfort or concern about having their consumption displayed outside of their apartments and visible to neighbors or passing people: "*It was indifferent… I was not at all concerned*" (House 1 Mother), although the consumption was only broadly exposed through the color of the landscape slice. When asked if they would feel the same way, if actual consumption data in form of numbers were displayed, instead of the forest representation, families again reported minor privacy concerns, but not strong opposition to it: *".. Of course there could be some privacy concerns… but I don't see any negative aspect with that*" (House 3 Father). It is important to note that asking families about a hypothetic scenario, their reaction could be different from their a priori answer once they could actually see the numeric information outside their house.

Particularly interesting reactions to the system came from kids. In one of the apartment blocks the children from the two families that participated in the study were friends, and in both cases they quickly became responsible for performing the task: "*He went there before going to school, he asked me for the consumption, and he went there … I have got the stack with the magnets, he chose it before going to school*" (Father, House 2), "*She went there before leaving to school. She got the magnet from the stack and we stopped in the mailbox before leaving*". (Mother House 1).

These statements suggest that competition arised between kids and that unspoken competition soon became apparent to the adults in the families: "*It seemed that they were competing with each other... every day in the morning she asked who spend more us or them… at the end she told me Who won?. It was more a competition between them, she was the one reminding me to check the message*" (House 1 Mother), "*Normally we don't spend without being necessary, but there was a day with higher consumption. He came to asked me why is that? I had to explain him that his mother was cleaning and so on*".(House 2 Father).

In the other two families (from a different block), the task was performed by the father, (House 3) and the mother (House 4). The positioning of the mailboxes meant that it was almost impossible for neighbors to ignore each other's consumption. This was confirmed during the interviews, 2 of them were actually curious to check: "*I got surprised in the end, our values were close (comparing with the neighbor), but of course since we are comparing savings the actual consumption values could be different)*" House 3 Father.

Fig. 3. Mailbox of one participant in the beginning (left) and in the end of the study (right)

6 Discussion and Conclusions

Eco-feedback is a promising technology to bring awareness to resource consumption. In this paper we reported an ongoing study looking at innovative ways to provide eco-feedback that leverage an emotional connection of people with their consumption and the long-term effects while at the same time making use of some playful strategies to raise awareness. WattsBurning on my mailbox provides an interesting strategy for families to display and share their energy consumption. Our preliminary results show that the task of placing magnets reflecting daily energy consumption outside of the houses was not considered hard or a burden by the families. When given the opportunity children become the gatekeepers for the magnets update, and function as probes and awareness aids of the adults about their consumption. We believe that this can provide opportunities for children oriented eco-feedback systems that are more playful and stimulate communication about routines and energy consumption patterns inside and across families. Additionally the fact that participants didn't show any concerns about publicly displaying their energy consumption suggests that there is space for exploration in public displaying and sharing of energy consumption. These promising results are inspiring the research team to develop the mailbox concept further and planning a larger deployment that could also lead to more statistically significant results in particular for the long-term effects in terms of consumption and awareness.

References

1. Wilson, A., Yost, P.: Buildings and the environment: the numbers. Environmental Building News 10 (2001)
2. Froehlich, J., Findlater, L., Landay, J.: The Design of Eco-Feedback Technology. In: Proc. CHI 2010 (2010)
3. Pierce, J., Odom, W., Blevis, E.: Energy Aware Dweling: a critical survey for interaction design for eco-visualizations. In: CHI 2008. ACM Press (2008)
4. Peschiera, G., Taylor, J.E., Siegel, J.A.: Response– relapse patterns of building occupant electricity consumption following exposure to personal, contextualized and occupant peer network utilization data. Energy and Buildings 42, 1329–1336 (2010)
5. Anderson, W.: White: Exploring consumer preferences for home energy display functionality. Report to the Energy Saving Trust. Bristol (2009)
6. Pierce, J., Paulos, E.: Materializing energy. In: Proc. DIS 2010 (2010)
7. Static! Increasing Energy Awareness, `http://www.tii.se/projects/static` (retrieved January 2, 2013)
8. Johnson, L., Broms, L., Katzeff, C.: Watt-Lite: energy statistics made tangible. In: DIS 2010 (2010)
9. Bllimeier, L., Baur, D.: Vision Energy, `http://www.seenby.com/lena-billmeier/vision-energy-fosil-fuels-sculpture` (retrieved December 28, 2012)
10. Holmes, T.G.: Eco-visualization: combining art and technology to reduce energy consumption. In: Conference on Creativity & Cognition (2007)
11. Nuage Vert: Nuage Vert Project (2008), `http://www.pixelache.ac/nuage-blog/` (retrieved December 28, 2012)
12. Wood, G., Newborough, M.: Energy use information transfer for intelligent homes: Enabling energy conservation with central and local displays (2005)
13. Foster, D., Lawson, S., Blythe, M., Cairns, P.: Wattsup?: motivating reductions in domestic energy consumption using social networks. In: NordiCHI (2010)
14. Filonik, D., Medland, R., Foth, M., Rittenbruch, M.: A customisable dashboard display for environmental performance visualisations. In: Berkovsky, S., Freyne, J. (eds.) PERSUASIVE 2013. LNCS, vol. 7822, pp. 51–62. Springer, Heidelberg (2013)
15. Odom, W., Pierce, J., Roedl, D.: Social Incentive & Eco-Visualization Displays: Toward Persuading Greater Change in Dormitory Communities. In: Proc. of Public and Situated Displays to Support Communities. Workshop at OZCHI 2008 (2008)
16. Bird, J., Rogers, Y.: The Pulse of Tidy Street: Measuring and Publicly Displaying Domestic Electricity Consumption. In: Workshop on Energy Awareness and Conservation through Pervasive Applications (2010)
17. Nisi, V., Nicoletti, D., Nisi, R., Nunes, N.: Beyond eco-feedback: using art and emotional attachment to express energy consumption. In: Proc. of C&C, pp. 381–382. ACM (2011)

Dynamic Spatial Positioning: Physical Collaboration around Interactive Table by Children in India

Izdihar Jamil[1], Kenton O'Hara[2], Mark Perry[3], Abhijit Karnik[1], Mark T. Marshall[1], Swathi Jha[4], Sanjay Gupta[4], and Sriram Subramanian[1]

[1] Department of Computer Science, University of Bristol, Bristol, UK
{jamil,karnik,sriram,mark}@cs.bris.ac.uk
[2] Microsoft Research, Cambridge, UK
oharakenton@gmail.com
[3] Brunel University, UK, Mobile Life, Interactive Institute, Sweden
mark.perry@brunel.ac.uk
[4] NIIT, Delhi, India
Swati.Jha@hiwel.in, SanjayG@tni.edu.in

Abstract. We present a study of how children demonstrate physicality during collaboration around interactive tables at school. Our results show that children tend to dynamically position themselves around the tabletop area to effect particular social outcomes. These movements around the tabletop allow them to enact coordination strategies in their social interactions with each other to manage their learning and task-based activities. Our analysis indicates the importance of understanding physical strategies and behaviours when designing and deploying interactive tables in classrooms. We discuss how the design of tabletops in school can embrace the extensibility of this technology, providing access for children to shape their own collaboration strategies during learning.

Keywords: Interaction techniques, tabletop, spatial formation, dynamic spatial position, collaborative learning, children and India.

1 Introduction

Since the introduction of interactive tabletops, we have seen a growing corpus of research exploring how collaborative activities and practices are organised. A key domain of interest in this space is concerned with understanding the collaborative practices of children around tabletops with a view to understand the potential value of these systems within educational contexts. Underpinning this is a body of educational research demonstrating the benefits of peer collaboration in learning activities, and consequently, peer collaboration is becoming an increasingly important concern in the educational experiences of children. What is central to these arguments is that learning is facilitated through encouraging all students to actively participate in exploring learning domains, discussing alternative ideas and perspectives, offering explanations, clarifications and justifications to each other and devising shared goals and plans of action, and thereby facilitating their understanding [1, 13, 28]. In light of these characteristics

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 141–158, 2013.

of collaborative learning, interactive tabletops have been argued to have various properties that promote the collaborative nature of these learning activities, for example, the ways that they organise groups around digital learning materials to encourage face to face discussion; how they enable simultaneous participation in collaborative learning, promote engagement and allow all students to share control and responsibility over the input and manipulation of learning materials; and how they make individual interactions at the tabletop visible to support shared awareness [6-8].

Central to these concerns are the ways that the communicative and collaborative practices around interactive tabletops are organised and managed. A significant strand of interactive tabletop research has sought to understand these practices and the factors that mediate them [5, 14, 21]. Such studies, for example have explored the extent to which participation is equally distributed across the group members [5]; the nature of talk in the collaboration [7]; ways in which tasks and activities are spatially partitioned and organised on the surface with a view for coordination and collaboration [5, 23]; the ways that object manipulation and orientation are used in collaborative understanding, communication, task organisation and so on [12, 16]. These have provided us with a rich understanding and characterisation of tabletop practices.

What is curious has been a tendency to refer to a rather static organisation of collaborators around the tabletop. While this is not an altogether unreasonable state of interaction to examine, we suggest that focusing on situations in which users are immobile underplays features of collaborative activity around tabletops, namely the bodily movement and dynamic positioning of actors around the tabletop for particular social and communicative effect. In highlighting these concerns, our aims are more than simply documenting another interesting feature of tabletop interaction. Rather, we want to highlight how these actions unfold and are produced in situ–with respect to social context of the group, the task and table–and what they achieve in the context of collaborative peer-learning tasks. We also want to explore how such a perspective might inform and refine our ways of thinking about key concepts in tabletop research, such as *territoriality*, *orientation* and the *spatial* organisation of action [14, 24].

To this end, we present a study of children engaged in a peer-learning task around interactive table in a school in Delhi, India. While it is of interest in adding this cultural setting to the predominantly western-based tabletop studies, it is not our intention to focus specifically on cross-cultural concerns. Of greater relevance is that the study is set in a real world classroom and school setting with the accompanying social dynamics of a real collaborative peer learning activity. Within this context, we explore the collaborative action of the peer learning groups and how these actions are produced, understood, and made visible in relation to the spatial and material arrangement of peers, and tabletop artefacts. In particular, we draw attention to a key feature of these practices–namely the dynamic use of spatial positioning around the tabletop. Before moving on to the details of the study, we begin with a look at the related literature to inform our understanding and ground our subsequent discussion.

2 Related Work

With the advances of commercially available systems such as the SMART Table and Microsoft PixelSense, the study of multi-touch interactive table use within a learning environment has become a growing area of interest. Looking beyond basic usability of such systems, this research has sought to understand the collaborative dynamics of children around digital tabletops and the benefits this might have for collaborative peer learning. One of the key motivating issues here has been on the equity of participation enabled by the multiple interaction access points of interactive surfaces. The shareable nature of these tabletops enables greater engagement by all group members by allowing simultaneous input contribution [20]. Studies have highlighted, for example, how this can have an effect on the nature of conversation during learning tasks leading to greater proportions of task and topic related utterances compared with utterances used for the purposes of task coordination [5, 7].

In Jamil et al. [7], children using interactive tabletops demonstrated task-based, reflective and teamwork related utterances during collaboration. Children went beyond discussing the presented topic and were seen reflecting on a higher level, relating the topic to knowledge and information from other sources (e.g. books and other classes). Such findings have particular significance for the success of collaborative peer learning that is the employment of particular types of talk, such as explanations and clarifications [28]. Similarly, the SIDES application [21] was able to motivate and support effective group skills between children with Asperger's Syndrome (who find working in groups difficult) in a four player cooperative game. Their findings include an increase in positive language use and a decrease in aggressive behaviour, building confidence in teamwork and offering a sense of participation and belonging.

Multi-touch systems themselves offer a different kind of platform for collaborative interaction compared to single touch systems. They allow users to spatially position themselves and simultaneously manipulate objects. In a single touch condition, only one person can manipulate an object at any time. This may not be a conducive environment for spatial movements to occur naturally: Harris et al. [10] suggest that single touch tends to result in a more turn-taking conversation that may indicate a turn-taking behaviour to follow these conversation patterns. Furthermore several researchers have shown that bodily movements are not seen when groups of students perform classification and sorting tasks around a non-digital tabletop [13,25]. One possible explanation could be that the experience of interacting with digital and physical objects differs: Marshall et al. [25] report that children in digital tabletop conditions were "more forceful" compared to the children in the non-digital conditions when blocking access to objects. This suggests that children may adopt different strategies when presented with different tabletop settings.

Other studies have focused on interactions on the tabletop and the embodied gestures around tabletop objects in collaborative activity. Visibility of these interactions in the face to face interaction of tabletop computing for example has been argued to promote an *awareness* of actions to others [6, 15]. This can be important in task coordination, but is also important in other features of peer learning such as consensus building, delegation of action, and production of explanations. Nacenta's work [23] shows how awareness in tabletop communication and collaboration is supported by three key mechanisms: *territoriality,* the division of the table surface into personal and shared regions for interaction (cf. [14]); *feedthrough,* namely the public visible manipulation of artefacts on the tabletop, for example Kruger et al's [14] notion of object *orientation* in communication; and *consequential communications* [25], the embodied actions produced by the hands, arms and bodies of group members as they collaborate. Let us consider these further.

The most significant work of territoriality in tabletop interaction can be seen in the seminal work of Scott et al. [24], in which the tabletop area is divided into four spaces: *personal, shared, group* and *storage*. The personal territory is that closest to a person and can be viewed as an extension of oneself [14, 24]. A participant can disengage with his or her group while exploring ideas before sharing it with the other members. Shared space occurs when a sub-group is formed and uses a particular area on the tabletop to direct their attention and discussion. The group space, usually located in the middle of the tabletop, is a communal area where all the members can jointly explore concepts together. The storage space serves as a holding area of unused resources and can be located anywhere around the tabletop. This partitioning influences how group members coordinate their activities when dealing with objects. Importantly, Scott et al. [35] make distinctions of territorialities along functional and spatial dimensions. In this sense, the exact size and position of these territories are socially constructed within the context of the task and setting. However, they also characterise these territories in terms of their relationship to the seating positions of the group. Implicit in their characterisation and representation of these spaces is a static set of seating arrangements. This presents strong implications for tabletop design to orient to territoriality in terms of its spatial rather than functional characterisation. While this is not an entirely unreasonable assumption, we want to argue that it at times ignores the potentially important role of dynamic bodily positioning around the tabletop in the construction of territorial arrangements of learning tasks. It is interesting to question how territoriality might be characterised when it is less bound and determined by fixed seating positions [14, 24, 27].

Within the context of these territorial practices are the *feedthrough* issues of orientation management–the degree to which objects on the tabletop are oriented to particular individuals or groups of collaborators: the orientation of an object on a tabletop towards an individual can confer certain territorial responsibilities that allow users to partition responsibility for particular features of the task. Such arrangements are bound to particular social protocols regarding who can manipulate particular objects on the tabletop within different territories. In this respect, interacting with objects by repositioning or re-orienting them is a negotiated concern among the collaborators. Within this context, orientation is actively used as a resource to achieve particular social effect and meaning (Kruger et al. [14]).

For Kruger et al., orientation of objects on tabletops serves three critical roles during collaboration: *comprehension of informatio*n, *coordination of actions* and *communication between members*. Thus reorienting an item is actively used for a variety of social and cognitive concerns, for example, to invite comment, hand over responsibility, initiate action, clarify intent, enable shared reading among collaborators, establish conceptual relationships among items, and so forth. What emerges from this is the situated production of meaning through orientation, that is meanings can be enormously varied. With such varied inscription of meaning through orientation, there are also potential points of conflict, in particular for exploratory and expressive learning tasks where spatial organisation of artefacts is also bound with conceptual understanding of the learning domain. Although useful, Kruger et al's explanation perhaps underplays the other means through which we may construct our relationship with objects on the tabletop and other members. Rather than creating meaning by orienting objects on the table, people can also simply reposition themselves with respect to the objects and other people: a feature of tabletop interaction that has been underplayed.

The work of Tang [27] provides us with more insight into spatial positioning. This highlights a number of spatial positions adopted by collaborating couples. These positions are used to achieve greater or lesser closeness of collaboration and avoid territorial conflict with couples adopting positions of closer proximity when working in a tightly coupled way. While Tang's work is broadly in line with the concerns of our paper, it presents a relatively abstracted characterisation. In this respect, we miss out on how the detailed use of bodily positioning and orientation at the tabletop operates as an embodied and situated practice, to be combined with other gestures and actions in relation to tabletop artefacts and collaborators, and the particular social meanings these practices achieve. The work of Yamashita et al. [29] provides us with a richer account of positioning and bodily orientation along these lines. This work is concerned in particular with how such action is organised around tabletops in distributed collaborative settings and how different video configurations and bodily views better enable the achievement of mutual orientation of awareness. As such it is less concerned with how such practices are achieved in collocated settings and in relation to the specifics of collaborative learning tasks. In the context of computer supported collaborative learning such accounts can be found in the work of Suzuki et al. [26]. Suzuki highlights the importance of seeing learning as participation in a culture of practice and the importance of designing to facilitate participation and interaction among learning groups. This is achieved through the use of gaze, body movement and gesture in relation to the computer artefacts. The study examines the interactional details of these practices in relation to group learning around a tangible computer interface. In this paper, we extend these concerns to consider how such interactional practices are organised in tabletop learning.

In informing our concerns we can also draw on some more general accounts and theories which while not specific to tabletops can nevertheless help frame our interpretations. Of particular significance here is Goodwin's [3] treatment of pointing as a situated and collaborative practice. For Goodwin, "pointing is constituted as a meaningful act through the mutual contextualization of a range of semiotic resources including at least, 1) a body visibly performing an act of pointing; 2) talk which both elaborates and is elaborated by the act of pointing; 3) the properties of the space that is the target of the point; 4) the orientation of relevant participants toward both each other and the space that is the locus of the point; and 5) the larger activity within which the act of pointing is embedded." Of significance is the collaborative nature of pointing and the multiple spaces and meaning making structures (for example artefacts, bodies of the actors) that are brought together to achieve it. Furthermore, this work starts to give us insights into some of the different purposes of these pointing (for example explaining vs directing action) and how is organised to achieve these effects. Kendon's work on gesture and spatial positioning [10] on the f-formation is also relevant. In this, when people communicate with each other, they mutually form f-formations (facing formations) that create mutually accessible transaction segments (o-spaces) between each participant. Attention and interaction are typically focused in this space with ongoing spatial and postural work being used to try and maintain these spaces. Indeed some other recent work in tabletop interaction has looked at spatial positioning based on Kendon's F-formation [10], most notably in Marshall et al. [17]. In contrast to the work we present in this paper, the concerns of the Marshall study were more with gross observable outcomes to see where social interactions did or did not work, rather than the detailed and ongoing production of social action. Our study builds on the concerns outlined here and explores the embodied and situated nature of collaborative learning practices around an interactive tabletop. In particular, it seeks to highlight the role and meaning of bodily positioning and orientation for the specific demands of collaborative learning, and this is reflected in the study that follows.

3 Study Description and Analysis

Participants. We recruited 51 pupils (between 11-13 years old) consisting of 27 females and 25 males from a local school in Delhi, India. The participants were divided into 11 groups of 4 to 6 pupils, a typical number for group-based classroom activities in these schools and fulfils the criteria for working in small groups [2]. This school is a private fee-paying school with mixed-sex students from different backgrounds. School children had access to a computer on regular basis. The medium of education is English and students talked in English during the entire study although they were not asked to speak in a specific language. The assignment of groups was conducted in consultation with the teachers in order to create groups of compatible ability levels.

Apparatus. We produced two custom-built FTIR tables [4]: 90cm x 90cm and 76cm high with a projection of 72cm x 48cm (NEC projector). Point Grey Dragonfly2 infrared cameras tracked user interactions. The task applications were created using Adobe Flash and Action Script 3. Tables were configured with two interaction techniques: (1) direct touch: a multi-touch platform where multiple users could interact directly with digital objects; (2) pantograph: finger movements in the 4 pantograph areas are amplified to create larger cursor movements on the surface, allowing digital objects to be reached without stretching across the table [18]. For larger groups, additional pantograph areas were added. Both tables were situated in the computer room at the school for the three days of the study. Chairs were available around the table.

Data Collection and Analysis. During sessions, children were exposed to both interaction techniques and learning activities, and after each session we switched the order of the conditions randomly. Video was used to record the physical and verbal behaviour of all the groups. Our analysis draws on a detailed examination of the video recordings and focuses on the collaborative achievement of the tasks. We describe the interaction details of how gestures, talk and action are produced, coordinated, made visible and understood with respect to the table and on screen objects. We used evidence-based interpretation to illustrate our findings and observations following the methodology of existing literature on interaction analysis [9]. The empirical data include sequence of vignettes depicting the children's interaction with digital objects.

Tasks and Techniques. Two collaborative learning activities were deployed: (1) spider diagram: similar to a mind map in which a topic is investigated and explored by visualizing associations and relationships between key concepts (Fig. 1a), and (2) classification: to classify and group twenty elements (images and concepts) of a topic according to category (Fig. 1b). Each category is represented by a square yellow box. The layout of the elements were scattered in a circle around the centre point of the surface providing equal access to the elements for each participant. The spider diagram and classification activities were based on "Photosynthesis" and "Animals", both topics in the Indian National Curriculum (further description in [7-8]): using different learning activities and interaction techniques were intended to add richness to the study, and ensure that these findings were not dependent on a particular educational topic or interaction method.

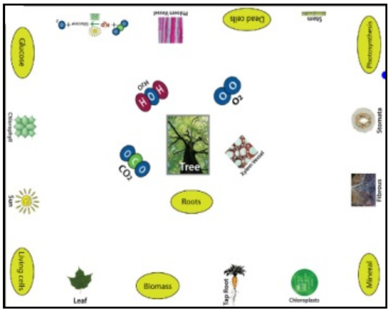


Fig. 1a. (L). Spider diagram (direct touch technique)

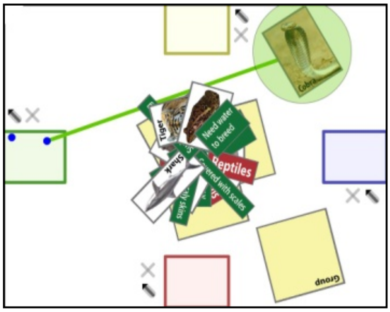
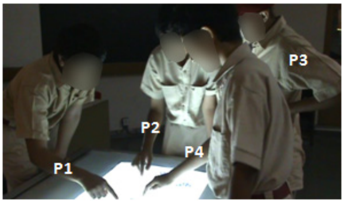


Fig. 1b. (R). Classification (pantograph technique; blue dots represent the user's fingers)

A number of features about the tasks and technology are worth raising here as they have particular implications for the interpretation of the spatially organised practices of interaction observed. First, the task at the outset is essentially orientation-free, i.e. the arrangement of items on the tabletop neither suggests nor commits the learning group to any particular spatial organisation. Second, the task itself has a spatial component to it- spatial relationships between tabletop items can and are used to explore conceptual relationships between items. While spatial structures are not imposed, spatial parameters such as position, mutual proximity and orientation can be used to achieve mutually understood conceptual organisation of the information. This is a standard feature of educational tasks that involve things such as classification or conceptual organisation of information. In the direct touch system, no interface elements favoured a particular spatial organisation of the group around the table. In the pantograph system, the elements at each side of the table could potentially encourage a particular spatial positioning of the students to act in controlling the system. With chairs positioned at the table the students were free to sit or stand and had room to move around. The dimensions of the table generally allowed the children to reach all parts of the table though for some this would require stretching to reach the furthest part. Notably there was nothing in the task, interface or hardware set up that was purposefully designed to induce the learners to move around the tabletop.

4 Results

In this first vignette, the group is discussing some initial classification for the structuring of *roots*, *stem* and *leaf.* P1 said: *"Put the roots next to the tree right."* At this moment, P1 gestures above the table to trace a path from where the "roots" piece is located to next to the *"tree"* piece (Fig. 2a).



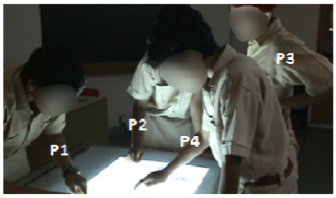


Fig. 2a (L). P1 pointing at roots above the tabletop **Fig. 2b (R).** P1 moves toward the corner

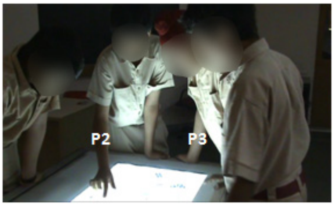
In doing this, P1 is suggesting a conceptual organisation on this part of the task, these pieces being within easy access of his current position. While he is doing this, P2 and P4 are both attempting to impose their own classification structures on the same pieces in the middle of the table, blocking each other's finger gestures as they compete to organise the pieces. P2 lets P4 complete his positioning, but then lifts P4's hand off the table to prevent any further interaction, allowing P2 to then bring another piece over to the emerging structure. P1 responds and says *"Don't you think we should put leaf there so we can put roots over here?"* [pointing to the appropriate locations on the tabletop to suggest a particular spatial and conceptual organisation]. P2 then suggests *"Put roots here."* (but moves the "roots" piece to a different location). Without saying anything, P1 moves further across the side of the table to position himself more in line with the newly positioned piece (Fig. 2b). In doing this P2 has built on P1's suggestion but through his action suggests a slight modification to the spatial structuring that they should use

A number of key things can be seen here. First, through this action and the continuation of work there is a non-verbalised demonstration of consensus with the structure offered by P2. Second, the continuation or work (and thereby the consensus demonstration) demands a particular bodily orientation with respect to the information artefacts on the table (Fig. 2b). We see a two handed manipulation as objects are brought into conceptual relationships with each other. Such a 2-handed manipulation would be awkward without moving either the artefacts or the body. Moving the artefacts was not a viable option since their spatial positioning was given meaning in the context of the task. To move them would have disrupted the emerging spatial structure, undoing progress and importantly, socially going against the proposed structure.

As P2 makes this manipulation, P2, P3 and P4 are focused on the other part of the table as they discuss other parts of the task. While there is distribution of the task activities here, they are not entirely separate and indeed there remain dependencies on the work being carried out by P1. In response to this, P2 breaks away from P3 and P4 and moves towards the corner of the table and reorients his body to be in line with P1 and the artefacts that P1 has been working on and asks *"P1 where did the leaf go?"*. As P2 asks the question he inspects the area where P1 has been working in order find the leaf and see what P1 has been doing. Both positioning and orientation are key at this point both for the cognitive and social demands of the task. Firstly P2 does not position himself right next to P1 but rather remains in a position somewhere between P1 and the other two participants–his engagement is temporary and maintains links across the subgroups. While in this corner position, his bodily alignment with the tabletop artefacts make the necessary comprehension and search easier. This is consistent with Kruger's articulation of orientation but as we again see in this instance, rotation of the tabletop structures is not possible as these are bound up with the spatial representation of concepts central to the learning activity at hand. As such orientation of the body is used instead. An additional feature of this orientation is to initiate a discussion with P1. Indeed what follows is that P2 leans in to move one of P1's pieces which then bring the work in line with the other three had been working on.

4.1 Orchestrating Participation

With the movement of P2 to the corner, we see an interesting behaviour arise from P3 and again this is one that is a key feature of such collaborative learning tasks. In moving to the corner, P2 actually creates some open space. Up until this point P1, P2 and P4 had been adopting positions along the sides of the table and as such were more actively involved both in the task both in terms of manipulating tabletop content and in terms of conversational contribution. P3 had been positioned slightly outside the primary f-formation made up by the other three participants and as such was struggling to get involved. With P2 moving and opening up a space between P2 and P3 (see the space created from Fig. 2b and 3a), P3 then moves closer into the table and spreads his arms across both sides of the table to establish and protect his position at the table, albeit a corner position (Fig. 3a). From this point, P3 is able to actively contribute to learning task through manipulation and conversation (Fig. 3b). Such behaviour is worthy of some comment here because it highlights the importance of being involved in the learning task and some potential difficulties of orchestrating such involvement with tabletop computing. Being actively involved in the task can be more engaging rather than passively watching. From a learning perspective, it allows the participant to explore their own conceptual contribution and have them questioned and critiqued by peers. But there are potential social concerns at play also in relation to participation in such learning tasks. For example, it can be important to demonstrate your attempts to contribute and demonstrate your intellectual grasp both to peers and potential assessors. While it is difficult to comment on the exact motivation in this particular context there are a number of things which are clear from the behaviour. First of all was how participation was inhibited by the particular form and size of the table that at times made it physically and socially difficult for some to manoeuvre into a position of participation. Second, the opportunistic spatial orchestration of participation by P3 demonstrates its importance to him in this learning context and for any number of the learning related reasons outlined above. As this is happening P4 carries out some manipulations in the middle of the table, then calls out, P4 said: *"The sun goes to the tree!"* [as he speaks he touches *Sun* and then touches *Tree*. P2 removes his hand from the middle of the tabletop. P4 then draws a line between those two keywords. P3 looks in the direction of the tabletop].



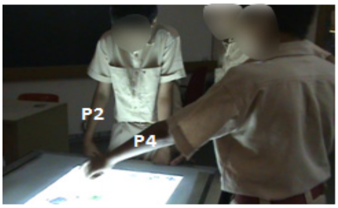


Fig. 3a (L). P3 moving closer to the tabletop **Fig. 3b (R).** P4 suggesting a new location

P4 said: *"And now branch is here… and then leaf and roots here!"* [P4 moves his right palm above the tabletop towards the corner right side of the tabletop, and the moved it again towards the corner of the left side of the table top and then towards the bottom of the tabletop. All the children then start to move objects from the middle of the tabletop towards the side of the tabletop leaving Tree in the middle. They were also scaling and rotating objects for visibility and orientation within this representational structure.] What is happening here is that a conceptual structure for the task is gradually coming into place as represented through an agreed upon spatial organisation. While the structure is not complete, by agreeing to the conceptual organisation it allows them to plan for how the task can be appropriately divided up. The movement of the objects to particular spatial locations is towards this end. It is only having got to this point through their spatial explorations that they can divide the task up in this way. Their arrangement of the objects in particular locations on the tabletop corresponds to this distribution of labour. Given this arrangement, what is curious at this point is how the enacted task distribution does not simply map onto their existing spatial positions. Rather they begin a sequence of delegation and volunteering to take on the subtasks that have been created.

This is initiated by P4 who says *"Okay, now, now I will do the roots stuff!"*. As we see in Fig 3b, P4 gestures towards an area of the table as he speaks and as he is doing this he is turning his body away to begin to move towards the area where the roots are (next to P2). As he is moving away he begins to delegate with an utterance and a gesture. He does not need to complete the delegation as the remaining boys jump into to try and get first digs on the other parts of the task. Both P2 and P3 simultaneously say *"I'll do the stem"*, P1 following with *"leaf"*. As these utterances are taking place P4 continues to move round towards P2's old position, while P2 and P3 both begin to move round the table to make way for P1 and also manoeuvre themselves into the correct position at the table to deal with the *"stem"* work (Fig. 4a). Of particular note is that P2 and P3 continue to move around until P2 is in prime position for *"stem"*, that is, P3 does not stand his ground but moves further along the table and gives into P2. Again, P3 is moved out of the way to a position where it is difficult for him to participate. Realising that he is spatially no longer in a position to actively participate. P2 moves round to stand between P1 and P4. From here he is able to actively get involved (Fig. 4b).

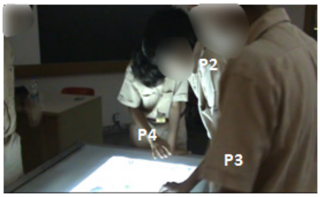


Fig. 4a (L). P2,P3 and P4 spatial positioning

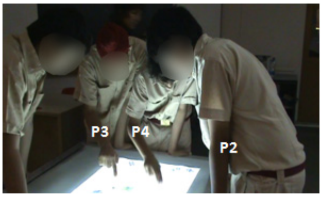


Fig. 4b (R). P3 shifting

What we can say about this intriguing sequence of actions is that the distribution of learning tasks is not one of simple spatial convenience whereby the learners are taking on the tasks that are closest to them. And indeed, the spatial representation of the conceptual relationships is manifest in order to be meaningful to the learning task as opposed to be organised around spatial convenience for interaction. Furthermore, these set of subtasks are evidently not considered equal and while it is difficult to ascertain the basis of such differentiation, it is clear from the behaviour that the participants consider it to be the case. This could be a question of certain parts of the task being more interesting to particular individuals, more suited to specialist knowledge of individuals, more intellectually demanding and therefore suited to certain individuals. Whatever these particular values, we can see too that certain social relations are being enacted through these spatial positioning and the ways in which peers displace or give way to the other peer collaborators. Again, we can only speculate as to the details of these but this does highlight is how such learning tasks in the real world do not exist in some social vacuum but within a whole range of existing social and intellectual relations between the peer learners that are an inherent part of these practices in real educational settings.

4.2 Orientation, Explanation and Clarification

Here, a different group is exploring the process of photosynthesis (initial position-Fig. 5a). P5 said: *"Look. This is the entire process of photosynthesis!"* [P5 points to a group of keywords (*O_2, CO_2, Phloem, Photosynthesis*) in the bottom lower right hand corner of the table (relative to him) using both of his index fingers (Fig. 5b)].

P5 said: *"All of this goes into photosynthesis!"*. He then traces an arc with both of his hands and turns his body at the same time to a second group of keywords

(*Mineral*, *Sun* and a few other non-visible keywords) at the bottom left corner of the table (relative to him). This arc and bodily movement are performed while looking at P3 and P4 with P1 and P2 following his movement. P5 then repeats the movement and said: *"This is all the same. This is photosynthesis!"*. P1 then answered: *"Oh my God! You're smart!"*. P5 then added: *"maybe bring all this here"* [he traces the same arc in reverse direction from his lower left corner to the lower right corner of the table]. P5 then continued: *"We need to move everything there!"*

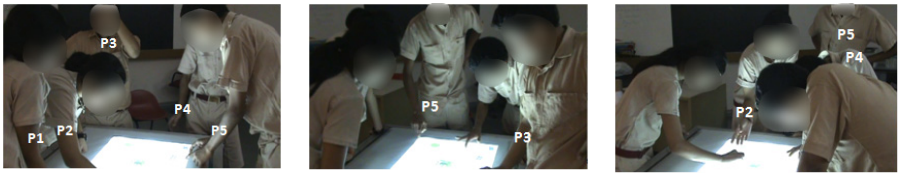


Fig. 5a (L). Initial position. **Fig. 5b (M).** P3 and P5 moving **Fig. 5c (R).** P2 shifting position

In this set of interactions we see that P5's conceptual breakthrough establishes a certain spatial order on the tabletop representations. The others, being impressed with this conceptual insight, then commit to further establishing the spatial relations this entails and begin to move the pieces to the location P5 has suggested and orient them in a way that is commensurate with the established grouping in the lower right hand corner of the table (relative to P5). This then commits the group to a new orientation. P3 who has moved to access the keyboard to the left of P5 then says to P5 *"Move that side"* and gently nudges him out of position, which continues the instruction. P5 initially resists but then moves round to P3's original position (Fig. 5b). This new position for P5 is also at the correct orientation for the spatial representation of the conceptual organisation suggested by P5. This allows him to continue the intellectual lead that he has set. P2 and P4 who are either side of P5 continue to work on the arrangement under the direction of P5 who is watching them. P5 stands back from the table with hands on hips, visibly indicating that he is letting the others work on the arrangement. At this point P4 goes to place the CO_2 incorrectly. P5 then takes action to intervene. He moves in closer to P4 and aligns himself with P4 just partially behind P4. As P5 moves he points to a location on the tabletop he says *"CO_2 is here, here!"* (Fig. 5b). What we see here is a key use of bodily positioning and orientation in collaborative learning for the purposes of explanation and direction. P5, as the person doing the explaining aligns himself both in relation to P4 (to whom he is explaining) and in relation to the spatial arrangement on the tabletop being referenced (Fig. 5b). Immediately after this sequence we see a related practice that is again of significance in relation to collaborative learning, namely questioning conceptual organisation and inviting clarification. Here P2 moves from her original side of the table and positions herself in front of P5 and P4 and facing the spatial representation being worked upon (Fig. 5c). P5 stands back to allow her into position.

P2 said: *"O_2 is …. Wouldn't O_2 be after photosynthesis?"* [she makes a gesture to indicate this suggested temporal relationship between O_2 and photosynthesis]. P5 then said: *"No."* Here then, P2, who is seeking an answer to a conceptual question, actively uses the alignment of herself with the spatial representation of concepts on the tabletop and with the person from which she is seeking clarification, P5.

4.3 Pantograph

We also noticed similar behaviour when the students used the pantograph technique. This is very peculiar as the main benefit of using the pantograph is the ability to reach far object without the need to move. Physical movement and positoning from one place to another is hardly necessary and yet the children demonstrated such. In the interaction below, we see two children from Group 4 move from their location to help another child resize a *Group Box*. The *Group Box* is an essential part of using the keywords as it provides a space where the children can classify and group the animals, collectively signalling their understanding of the task.

P2 said: "*You just made it bigger!" [referring to the Group Box]*. Immediately after he does this, P1 hovers his left hand just above P4's pantograph (where P4's hand was located). P1 then moved closer towards P4's location and at the same time moved his right index finger from his pantograph towards P4's pantograph. P4 then released his touch on his pantograph. P1 touched P4's pantograph, highlighted the Group Box and started to scale it down. However, after some 20 seconds or of P1 scaling the Group Box without much success, the Group Box still had its original size.

P2 said: *"Make it smaller! Make it smaller!"*. At this point, P3 stood up and moved towards P4's pantograph area. P3 stretched his left hand and touched his index finger on P4's pantograph area, which he used to scale down the size of the object. P2 then said: *"He's a genius!"*. P1 then moved closer to P4's position, touched P4's pantograph and further scaled down the object. What is apparent from the interaction above is that P1 and P3 both moved position and used P4's pantograph to: 1) help and assist P4 and P1 to perform an action (in this case, scaling down a *Group Box*), and 2) to stop P4 from doing something that is visibly seen to go against the group's decision (i.e. making the *Group Box* bigger). What is curious in this interaction is that neither P1 or P3 used their own pantograph in order to resize the *Group Box* despite the available facilities to do so, but move and position themselves near P4's pantograph area and subsequently use this to perform the action. Given that this was previously his personal working area, it is notable that P4 showed little resistance to this.

We speculate a number of possible reasons for this behaviour: 1) the children perceive territoriality and ownership of an object based on the object's location. In this case, the Group Box is located within P4's personal space, hence P1's and P3's movement towards that area, 2) as territoriality of Group Box was mapped to belong to P4, perhaps using P1's and P3's pantograph would considered to be rude and an act of intrusion of object ownership, 3) the orientation of objects may determine the physical positioning of a user, in this case the position of *Group Box* was orientated towards P4, and 4) the children were demonstrating meaningful social attributes (for example teamwork), in particular, demonstrating the case of being supportive and helpful with each other, even though there was no verbal request for help from P4 to concur with the decision to resize the object. The ability of the pantograph to address physical 'reach' of digital objects outside of a user's arm length would seem to make such movements unnecessary, and given their experience with the pantograph by this stage in the study (and the fact that they use P4s pantograph), it cannot simply be that they are struggling with the interface. Clearly, spatial physical movement carries value above and beyond the physical necessities of interface.

5 Discussion

Through the fieldwork, we have begun to highlight the situated and embodied practices enacted by groups of students from Delhi performing collaborative peer learning activities on an interactive tabletop. In particular, we have focused on the ways that these learners employ dynamic bodily orientation and positioning around the tabletop for particular social, cognitive and collaborative effect in the context of these collaborative learning tasks. Such actions are made meaningful in the particular context of learning and are collaboratively organised in relation to the other learners, the tabletop and the task based artefacts on the tabletop. While such bodily arrangements and positioning are a resource in any collaborative activity, what is of interest is the particular ways that they are manifest in the context of these tabletop learning activities, why they are important in these activities, and the particular implications this has for our theories and design of tabletop interaction.

Central to the arguments we wish to make here is the cognitive use of spatial representations that is fundamental to such exploratory learning activities and tasks [11]. Within these tasks, domain knowledge comes to be explored and represented through the spatial relationships between the tabletop entities. Orientation, position and relational proximity between these different entities are all parameters through which the learners collaboratively come to represent key conceptual relationships in the knowledge domain both as tentative propositions to discuss and ones to which the learners eventually commit to as more static representational structures. Significant here is that such spatial representations of knowledge are emergent features of the learning task as it proceeds towards some form of resolution. Such an additional layer of spatial meaning then is another concern- embodied actions of the learner must be organised and that takes us beyond the key theoretical apparatus of existing tabletop research. For example, while territoriality remains an important and demonstrated concern within tabletop collaboration that has its expression in spatial relationships among actors and artefacts, this expression is made considerably more nuanced as actors orient to the additional layers of spatial concern in these tasks.

Similarly, as we have seen in the findings, the spatial configurations of actors around the tabletop requires a richer account than simply the management of relative closeness or looseness of collaborative coupling, though this is indeed important. With regards to Kruger et al. [14] articulation of object orientation in tabletop collaboration, our findings would appear to align with a number of the theoretical concerns of this work. That is in the role of tabletop object orientation in comprehension, communication and coordination work. But as we have seen in these learning tasks, the orientation and positioning of objects takes on additional significance in the form of the spatial representation of knowledge. In this respect the orientation of objects is but one resource that can be used to achieve different forms of comprehension, communication and collaboration work but is one that is not always available as objects gradually adopt their positions in the spatial representation of knowledge. As such, bodily positioning and orientation with respect to the objects becomes an additional resource through which this kind of work is achieved. Such a resource can be especially important in the management of orientation and positioning in the real world learning contexts here where the larger group sizes configured around the tabletop present more complex orientational and positional dependencies between the actors and artefacts which need to be managed. With this in mind, then, let us consider some of the more specific ways in findings that this use of bodily orientation and positioning around the tabletop was used and how these have particular significance in the context of the collaborative learning activities:

1) In the data, a very pragmatic concern arises from the need to manipulate multiple objects into meaningful spatial relations with each other to represent conceptual relationships. When conducted as a bimanual activity, this entails certain physical demands on bodily orientation and positioning in order to bring objects together.

2) Bodily orientation and positioning was used to facilitate the interpretation of conceptual relationships being expressed spatially by the other learners. This involved alignment with the orientation of these particular spatial structures, but to do so in ways that also maintained orientation to other structural features of the representation. This allowed interpretation of sub-parts of the task in the context of the whole.

3) Movement and positioning was used to initiate enquiry into particular conceptual representations, one of the significant features and benefits of collaborative learning. In doing this, bodily orientation and positioning is organised both in relation to the information and the other learners involved. This also sets up the possibilities and permissions for subsequent manipulation of the spatial representations created by others allowing them to build and refine the conceptual work being performed.

4) Positioning and orientation are used to orchestrate and enable participation. Of significance here is that with the larger number of learners in these real world learning groups, some learners end up in positions around the table where they are disadvantaged in their access to the active transactional spaces in play. In such circumstances, they opportunistically seek positional openings as they arise in the positional configurations of the group. Such participation is key to learning in these groups allowing individuals to explore their own conceptual understandings. Likewise, such visible attempts to participate are not only important in demonstrating contribution to the group but in the context of real world educational settings may be important to assessment. Furthermore, such opportunities to participate and contribute are emergent in the context of the task itself. Only as the task progresses do it become apparent and mutually agreed how the task can be divided up into subtasks. Here again, positional manoeuvring can be used to orchestrate participation and ensure influence over the emerging transactional spaces. In such manoeuvring, it is not simply the characteristics of the task and artefacts that are brought to bear on action, but also the pre-existing social relations of the group members from the broader school experience itself that brought forth a whole host of existing social relations among the learners.

5) While positioning and orientation enables participation, this goes beyond the interaction technique settings, i.e. the children violate the “supposed” working pattern when using the pantograph technique. Thus, although the pantograph allowed equitable access to objects within the tabletop area [18], students can be seen using other people’s pantographs. Of significance here is the value that the orientation and positioning brings forth during the activity. Physical movement allows the children to continue with the activity without interrupting the existing spatial flow and relationship that is crucial in their learning. Perhaps positioning themselves at another member’s pantograph also ties in with territoriality and orientation notions above. In one hand, the children wanted to help another member as part of demonstrating group cohesiveness, while on the other hand they were aware that the object does not belong to them (i.e. it was not orientated towards them and not located within their personal space [14, 22]. They respected the notion of ownership and thus moved themselves accordingly in order to gain ownership in the process of helping others.

6) Finally, we see the important use of physical orientation and repositioning in relation to key features of collaborative learning, namely to direct and explain, and to invite clarification of proposed conceptual structuring. Notable here was that not all participants have equal status within the groups at different points in the task. Rather at any time, certain individuals may have key insights that emerge that lead to particular spatial representations and around which the group spatially converge and achieve consensus. This can demand the intellectual lead be positioned at the dominant positioning in the transactional space. Key here is that in directing and explaining the concepts to others, the person directing or explaining will align themselves with both the spatial representation and the individual explainees so as to shape the explanatory gestures and talk to the perspective of the explainee. In a similar vein, we see those inviting explanation and clarification adopt similar mutual alignment with the representation and explainer to allow both articulation of their misconceptions and appropriate receipt of explanation.

In evidence here then is the situated and collaborative achievement of these significant features of collaborative learning practices. In such practices, and to paraphrase Goodwin [3], bodily orientation and positioning around the tabletop are constituted as meaningful acts through the mutual contextualization of the semiotic resources to hand and involves the orientation of participants to each other and the representational space of artefacts. What we want to argue here, is that such actions are an important part of a repertoire of resources available for such meaning-making much in the same way that, for example, Kruger et al’s [14] orientation of tabletop artefacts might be regarded. In many aspects of these spatial and exploratory learning tasks they offer some potentially significant benefits when other resources (such as object manipulation) may be constrained or unavailable.

6 Design Implications and Conclusion

As well as the theoretical implications we have discussed above, there are particular implications for how we approach the design of tabletop interactions for these settings in ways that enable rather than constrain these important actions. For example, some applications and interaction mechanisms impose certain location dependencies on the users that may make them less flexible in their use of spatial positioning and orientation around the tabletop. For example, the idea of user identification and object ownership at one designated location (i.e. A stays at location X throughout the task with her personalized tools) or the capability of restricting user input (such as the features available in [21]) may impose restrictions along these lines. Using a fixed and digital control bar as suggested by Olson et al. [19] may impose constraints for similar reasons. What may be important is for applications to be fluid, dynamic and moveable. Mobile storage bins [29] and adaptive personal territories [12] may provide some interesting solutions here where learners have the capability of grouping objects (in our case keywords and pictures) and moving them around as they position themselves around the tabletop. But there remain questions over such whether storage bin/personal territories move automatically as the children move or whether they should be moved manually by their owner. In raising these our intention here is not to be overly prescriptive in the design suggestions nor to suggest that location-bound interaction mechanisms on the tabletop entirely prevent the use of bodily orientation and positioning as a resource (indeed, even in the positional constraints imposed by the Pantograph we see some deployment of these activities). Rather, our intention is to highlight the importance of these action resources in collaborative learning around tabletops and prompt reflection on how particular interaction mechanisms may constrain or hinder their enactment.

Acknowledgements. We would like to thank KPT (810502055410) and UPM (T01839) in Malaysia for partially sponsoring this research. A huge appreciation also goes to the Bristol Interaction and Graphics Group, NIIT, all of the teachers, staff and participants.

References

1. Eisenberg, A.R., Garvey, C.: Children's use of verbal strategies in resolving conflicts. Discourse Processes 4, 149–170 (1981)
2. Engleberg, I., Wynn, D.: Working in Groups. Communication Principles and Strategies. Houghton Mifflin Company (2000)
3. Goodwin, C.: Pointing as Situated Practise. In: Kita, S. (ed.) Pointing: Where Language, Culture and Cognition Meet, pp. 217–241. Lawrence Erlbaum (2003)
4. Han, J.Y.: Low-cost multi-touch sensing through frustrated total internal reflection. In: Proc. of UIST, pp. 115–118. ACM (2005)
5. Harris, A., Rick, J., Bonnett, V., Yuill, N., Fleck, R., Marshall, P., Rogers, Y.: Around the table: are multiple-touch surfaces better than single-touch for children's collaborative interactions? In: Proc. of CSCL, pp. 335–344 (2009)
6. Hornecker, E., Marshall, P., Dalton, N.S., Rogers, Y.: Collaboration and interference: Awareness with Mice or Touch Input. In: Proc. of CSCW, pp. 167–176. ACM (2008)
7. Jamil, I., O'Hara, K., Perry, M., Karnik, A., Subramanian, S.: The Effects of Interaction Techniques on Talk Patterns in Collaborative Peer Learning around Interactive Tables. In: Proc. of CHI, pp. 3043–3052. ACM (2011)
8. Jamil, I., Perry, M., O'Hara, K., Karnik, A., Marshall, M.T., Jha, S., Gupta, S., Subramanian, S.: Group Interaction on Interactive Multi-touch Tables by Children in India. In: Proc. of IDC, pp. 224–227. ACM (2010)

9. Jordan, B., Henderson, A.: Interaction Analysis: Foundations and Practice. The Journal of Learning Sciences 4, 39–103 (1995)
10. Kendon, A.: Conducting interaction. Patterns of behavior in focused encounters. Cambridge University Press (1990)
11. Kirschner, F., Paas, F., Kirschner, P.A.: A Cognitive Load Approach to Collaborative Learning: United Brains for Complex Tasks. Ed. Psy. 21, 31–42 (2009)
12. Klinkhammer, D., Nitsche, M., Specht, M., Reiterer, H.: Adaptive personal territories for co-located tabletop interaction in a museum setting. In: Proc. of ITS, pp. 107–110. ACM (2011)
13. Kruger, A.C., Tomasello, M.: Transactive discussions with peers and adults. Developmental Psychology 22, 681–685 (1986)
14. Kruger, R., Carpendale, S., Scott, S.D., Greenberg, S.: Roles of Orientation in Tabletop Collaboration: Comprehension, Coordination and Communication. In: Proc. of CSCW, pp. 501–537. ACM (2004)
15. Marshall, P., Fleck, R., Harris, A., Rick, J., Hornecker, E., Rogers, Y., Yuill, N., Dalton, N.S.: Fighting for Control: Children's Embodied Interactions When Using Physical and Digital Representations. In: CHI, pp. 2149–2152. ACM (2009)
16. Marshall, P., Morris, R., Rogers, Y., Kreitmayer, S., Davies, M.: Rethinking 'Multi-User': An In-the-Wild Study of how Groups Approach a Walk-Up-and-Use Tabletop Interface. In: Proc. of CHI, pp. 3033–3042. ACM (2011)
17. Marshall, P., Rogers, Y., Pantidi, N.: Using F-formations to Analyse Spatial Patterns of Interaction in Physical Environments. In: Proc. of CSCW, pp. 445–454. ACM (2011)
18. Nacenta, M.A., Pinelle, D., Stuckel, D., Gutwin, C.: The effects of interaction technique on coordination in tabletop groupware. In: Proc. of GI, pp. 191–198. ACM (2007)
19. Olson, I.C., Leong, Z.A., Wilensky, U., Horn, M.S.: "It's just a toolbar!" Using Tangibles to Help Children Manage Conflict Around a Multi-Touch Tabletop. In: Proc. of TEI, pp. 29–36. ACM (2011)
20. Pawar, U.S., Pal, J., Gupta, R., Toyama, K.: Multiple mice for retention tasks in disadvantaged schools. In: Proc. of CHI, pp. 1581–1590. ACM (2007)
21. Piper, A.M., O'Brien, E., Morris, M.R., Winograd, T.: SIDES: A cooperative tabletop computer game for social skills development. In: Proc. of CSCW, pp. 1–10. ACM (2006)
22. Rick, J., Harris, A., Marshall, P., Fleck, R., Yuill, N., Rogers, Y.: Children designing together on a multi-touch tabletop: an analysis of spatial orientation and user interactions. In: Proc. of IDC, pp. 106–114. ACM (2009)
23. Rogers, Y., Lim, Y.-K., Hazlewood, W., Marshall, P.: Equal Opportunities: Do Shareable Interfaces Promote More Group Participation Than Single User Displays? Human-Computer Interaction 24, 79–116 (2009)
24. Scott, S., Sheelagh, M., Carpendale, T., Inkpen, K.: Territoriality in collaborative tabletop workspaces. In: Proc. of CSCW, pp. 294–303. ACM Press (2004)
25. Segal, L.: Designing Team Workstations: The Choreography of Teamwork. In: Local Applications of the Ecological Approach to Human-Machine Systems (1995)
26. Suzuki, H., Kato, H.: Interaction-level support for collaborative learning: AlgoBlock-an open programming language. In: Proc. of CSCL, pp. 349–355 (1995)
27. Tang, A., Tory, M., Po, B., Neumann, P., Carpendale, S.: Collaborative coupling over tabletop displays. In: Proc. of CHI, pp. 1181–1190. ACM (2006)
28. Teasley, S.: The Role of Talk in Children's Peer Collaborations. Developmental Psychology 1, 207–220 (1995)
29. Yamashita, N., Kuzuoka, H., Hirata, K., Aoyagi, S., Shirai, Y.: Supporting fluid tabletop collaboration across distances. In: Proc. of CHI, pp. 2827–2836. ACM (2011)

Dynamic Tangible User Interface Palettes

Martin Spindler[1], Victor Cheung[2], and Raimund Dachselt[3]

[1] User Interface & Software Engineering Group, University of Magdeburg, Germany
[2] Collaborative Systems Laboratory, University of Waterloo, Canada
[3] Interactive Media Lab Dresden, Technische Universität Dresden, Germany
spindler@ovgu.de, victor.cheung@uwaterloo.ca, dachselt@acm.org

Abstract. Graphics editors often suffer from a large number of tool palettes that compete with valuable document space. To address this problem and to bring back physical affordances similar to a painter's palette, we propose to augment a digital tabletop with spatially tracked handheld displays. These displays are dynamically updated depending on their spatial location. We introduce the concept of spatial Work Zones that take up distinct 3D regions above the table surface and serve as physical containers for digital content that is organized as stacks of horizontal layers. Spatial Work Zones are represented either by physical objects or on-screen on the tabletop. Associated layers can be explored fluently by entering a spatial Work Zone with a handheld display. This provides quick access and seamless changes between tools and parts of the document that are instantly functional, i.e., ready to be used by a digital pen. We discuss several use cases illustrating our techniques and setting them into context with previous systems. Early user feedback indicates that combining dynamic GUI functionality with the physicality of spatially tracked handheld displays is promising and can be generalized beyond graphics editing.

Keywords: Tangible user interface palettes, spatial Work Zones, tabletop displays, dynamic pen heads, spatial management.

1 Introduction

Graphical User Interfaces (GUI) have been commonly used as the medium of interaction with computer applications due to their iconic appearance and ease of control. With GUIs, tools are often organized as tool *palettes* that group similar functionality within separate windows floating on top of the document. One drawback, however, is that these palettes take up the very same display space as the document does. This can be a major issue when applications support a large number of different tools. A typical example for this are graphics editors, such as Adobe Photoshop and Illustrator. Due to the wide spectrum of functionality – e.g., navigation, layer handling, system control, selection management, brush and color selection, etc. – many tool palettes are often displayed simultaneously, cluttering valuable display space. A common way to approach this problem is to rely on more transient interfaces, such as context menus, menus from the menu bar, or see-through widgets [5] that do not permanently occlude

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 159–176, 2013.

objects of interest. Unfortunately, this still forces users to spend considerable time and effort on organizing and navigating through tool hierarchies, thus taking away mental resources from their actual goals. One naïve solution to this problem is to provide more room for the interface and document by using larger displays, such as digital tabletops. Yet, the underlying conflict still remains: the strife between user interface and document over the very same display space.

Meanwhile, recent developments show that large interactive displays are gaining popularity in a variety of application domains. This is due to the fact that they fuse input and output in the same space, thus allowing for a more natural and intuitive interaction by using the concepts of direct pointing and manipulation. This is also supported by the ample size of such displays that provides an adequate environment for co-located collaborations. Previously, several projects have focused on the design and implementation of such displays in different configurations, either horizontally (e.g., interactive tables [8]), vertically (e.g., digital whiteboards [9]), or tilted (e.g., sketching table [25]). Others have created mixed display environments that benefit users with the best of each individual setup, for instance, by using display walls for presentation tasks, tabletops for collaborative work on shared documents, and smaller mobile displays for personal tasks [19, 20].

Apparently, content transfer between devices plays a vital role in such mixed display environments. For that, a variety of techniques have been presented aiming at an almost ubiquitous handling of content between the "digital" and the "analog" world, e.g., by using paper-like digital interfaces [13, 15]. Others, in turn, have demonstrated how the spatial position and orientation of such paper-like displays can directly be utilized for interaction purposes – used, for example, for zooming in/out details of large image documents that are displayed on a tabletop by moving a handheld digital paper screen up and down [29]. Most of these projects have focused on the problem of how to interact with and move *data* between different workspaces. Yet there is little work on investigating how spatial movements and arrangements of *tools* facilitate the work practice, which is the goal of our work.

To address this problem, we build on the benefits of tangible displays in a tabletop environment that extend the interactive space to dynamically updated and spatially tracked handheld displays. Our contributions are: (1) the physical separation of document and tools by using digital pen-enabled paper-like handheld displays, (2) the use of the 3D space above the table for interaction by introducing spatial Work Zones that provide quick and easy access to either tools or the document as well as fluid switches between the two of them, and (3) the demonstration of their usefulness by several use cases and a prototypic implementation.

In the remainder of the paper, we first review related work and identify a gap between the "analog" and "digital" world by using the example of graphics editing. Next, we propose our concept of Tangible User Interface Palettes (TUIP) that aims at closing this gap. This includes the description of the design space and our concept of spatial Work Zones. We then present several use cases that demonstrate how TUIP principles can help users organize and work with complex graphical user interfaces, as illustrated with a graphics editor application. After a brief walk through the technical setup, we report early user feedback and finish with some concluding remarks.

2 Related Work

Our work is based on a variety of previous research that we extend and combine within a single system. This includes conceptual as well as technical aspects.

2.1 Tangible User Interfaces: Physical Separation of Tool and Document Space

Merging the digital world with the physical world is the vision of ubiquitous computing, as defined by Weiser [35]. This concept was adapted by Ishii's and Ullmer's tangible user interfaces (TUIs) [17], where interaction with digital information is provided through physical manipulation of real-world objects (tangible objects). TUIs have been used to cope with the conflict between document and tools over the same display space by supplying additional multi-purpose tangible objects that users can manipulate with their hands, but without involving the document view. One example for this is using real brushes in different sizes and shapes, such as demonstrated with IntuPaint [34], though this still relies on virtual color widgets on the main screen. By adding interactive components to a digital pen, such as multi-touch capability to a pen's barrel, as demonstrated by Song et al. [28], users can switch between different interaction modes or drawing styles. I/O brush [27] goes a step further and lets children draw with digital colors and textures that they pick up from the real world.

Tangible GUIs aim at bridging the familiarity of conventional GUIs and the rich affordances of tangible objects. For example, controls can easily be organized and passed around. In addition, the main display is less cluttered, as most of the GUI components can be "detached" from the main display. In Phidgets [10], Greenberg and Fitchett demonstrated the support for user-adaptation of physical UI components using wired controls. The VoodooSketch project by Block et al. [6] extended a tabletop environment with printed paper-based GUI palettes that were physically separated from the main display, and supported live sketching of custom-shaped controls and their configuration by handwritten labels or prefabricated components (e.g., buttons and sliders) on physical palettes. However, once created, these interface palettes remained static until they were physically altered or thrown away.

In contrast, our work is based on digital paper-like displays, where digital content is dynamically projected onto physical palettes, and thus can easily be adapted and manipulated with immediate visual feedback. In this way, our project shares principles from the Paper Windows project by Holman et at. [15] that addresses the capturing of physical affordances of paper in a digital world.

2.2 Projective Display Technology

Compared with active displays, e.g., smart phones and tablets, projective (paper-like) displays are typically more lightweight, customizable, almost "tech-free", and are inexpensive to reproduce once the system is running with at least one display. Most projective display systems follow the same technical approach. First, the spatial positions of one or more handheld projection screens are determined, e.g., optically by using infrared (IR) cameras and IR-reflective markers [15]. Then content is projected onto the displays by a stationary projector that either resides above or in front of the workspace (e.g. ceiling-mounted [15, 24, 29]) or even below a tabletop surface, such as demonstrated in Second Light [18] and UlteriorScape [20]. One of the great potentials of projective displays is its support for digital pen and paper technology (e.g., Anoto-based [1]), which further helps merge ”analog” with ”digital” content, making it an ideal technical base for ubiquitous computing. On the technical side, our TUIP prototype utilizes projective paper-like display technology, in particular the one used and presented in our own previous work PaperLens [29].

2.3 Tangible Magic Lens Interaction (in a Tabletop Environment)

Inspired by the notion of see-through interfaces [5], tangible user interfaces can also be spatially tracked handheld displays (Tangible Displays) that serve as Tangible Magic Lenses into a virtual world. Tangible displays have successfully been used in tabletop environments, for example, in the metaDESK [33] project, where users can explore a virtual 3D campus by moving an arm-mounted TFT-display above a tabletop showing a digital 2D map of the campus. In PaperLens [29], we demonstrated how such Tangible Magic Lenses (handheld displays) can be used to navigate through individual parts of a large image document that is displayed on the tabletop. Here, lifting and lowering a handheld display correlates with instant zooming, whereas moving a display horizontally results in panning. In PaperLens, all spatial interaction was implicit, meaning that any positional change of a display triggered immediate updates on its screen. With Tangible Views [30], we introduced the concept of “freezing” that allowed for actively locking the screen content by holding a button, thus preventing any visual updates as long as the button was pushed. Yet, implicit spatial input based interaction still remained the default. For TUIP, we use similar techniques for document navigation. However, to better meet the requirements of graphics editor applications, where unintended navigation can be very disturbing, users must explicitly activate any spatially based navigation by holding and pushing a physical button attached to a handheld display.

2.4 Proxemics Interactions

The field of proxemics interactions originates from studies of interpersonal behavior in relation to physical distances. It has been adapted and widely studied in human-computer interaction as a means of providing appropriate responses based on user’s proximity [3, 14]. Another adaptation is to make use of the spatial relation between tangibles and adjust their behavior. In their initial work of the Smart-Its project, Holmquist et al. [16] proposed context proximity which connects artifacts based on physical distances and user actions (e.g. shaking the artifacts). Kray et al. [22] investigated the use of spatial regions around mobile devices (tracked using on-screen markers) for content sharing, and demonstrated their benefits on group coordination and social processes. In our work, we combine proxemics interaction principles with tangible display interaction by introducing spatial Work Zones, which are distinct spatial regions above the tabletop with a certain meaning associated to them.

2.5 Multi-layer Stack Interaction with Tangible Magic Lenses

One particular spatial interaction style for tangible displays is multi-layer interaction, such as was proposed and demonstrated by us with PaperLens [29]. The principle idea behind multi-layer interaction is to arrange virtual 2D layers in a vertical stack within the physical 3D space, with each layer consuming a distinct height of the volume. Such multi-layer stacks can be explored by moving handheld displays through it. Vertical movements of a display along the Z-axis allow for selecting a layer, while holding a display at a certain height allows for viewing or working with a particular layer. At the same time, moving the display horizontally allows for exploring a specific layer. In [31], we conducted an extensive user study and derived specific design guidelines for this style of interaction. Later, in LightSpace [36], Wilson and Benko projected spatial vertical menus directly onto the hand of users, thus suggesting a similar way of interaction by revealing menu options when the hand is being moved up and down. For TUIP, we use multi-layer stack interaction principles for organizing related tool palettes or documents within spatial Work Zones.

3 Identifying the Gaps between Two Opposite Worlds

We start our exploration by comparing the benefits and weaknesses of tool handling in two opposite worlds: the real (analog) world and the digital world. By using the example of graphics editing, we will identify significant gaps between both worlds. Narrowing these gaps was a major motivation for working on TUIP.

3.1 Graphics Editing in the Analog World

Prior to the digital age, painters and graphics artists used to work with real-world tools and painting mediums (e.g., brushes, color palettes and canvases), which were physically separated from each other and could be spatially arranged, grouped, manipulated, and combined within the physical 3D world in a very natural and personalized way. This style of interaction implicitly leverages the whole spectrum of spatial arrangement, which has proven to simplify choice, perception, and even internal computation [21]. After all, in having a body we are spatially located beings. We must always face some direction and therefore have only certain objects in view. Thus, managing the spatial arrangement of items around us is not circumstantial – it is an integral part of the way we think, plan and work. In this spirit, artists freely setup and configure their working environment to fit the task at hand, e.g., by defining and using specific work zones on the desk. This gives them not only fast access to all necessary tools, but also requires them to memorize less – by actively increasing their understanding of the spatial organization in the outside world. This is supported by human perception that can handle a high amount of information in the periphery, even if the focus is on a particular detail [32].

As a downside, working with the analog world is often limited by physical constraints of tools and materials. For example, existing drawings are hard to reproduce and supplies will run out or deteriorate after prolonged use.

3.2 Graphics Editing in the Digital World

In the digital world, graphics artists can reuse, share, modify, and combine digital content very efficiently by using digital tools that simulate analog techniques (e.g. airbrushes, paint brushes, erasers). This increases productivity and even complements the toolbox with novel styles, techniques and filters that were not available before in the analog world. Beyond that, many physical constraints are removed in the digital world. For example, colors never run out and the drawing space is virtually limitless. This is reflected by the extensive set of tools offered by common desktop graphics editors like Adobe Photoshop, Adobe Illustrator and GIMP.

One major disadvantage, however, is the restriction to a two-dimensional (2D) workspace: the digital screen that provides only a spatially constrained peephole into an otherwise large and complex world. This effectively takes away the advantages of spatial arrangement that users are familiar with from the analog world. Two challenges go hand in hand with this: tool management and document navigation.

Challenge 1: Tool Management. The immense amount of digital functionality increases the complexity of tool usage. This results in valuable screen space being cluttered by tools, leaving less space for the actual document. Common approaches to soften (but not solving) this problem include keyboard shortcuts, tool presets, context menus (e.g., Pie/Marking Menus [23]), and menus from the menu bar, i.e., transient interfaces that do not permanently occlude objects of interest – thus saving valuable screen space. Some of these techniques were adapted to multi-touch displays, e.g., Finger-count Menus [2] or Bimanual Marking Menus [11]. Most of these interfaces group related tools in form of nested windows that float above the document. This helps users to remember where a particular tool can be found, but also requires them to spend considerable time and effort on organizing and arranging UI-components.

Challenge 2: Document Navigation. By looking through the eye of a spatially constrained display, the challenge of viewing large graphics documents at various levels of detail should not be underestimated. Users often have to switch between different views that can either depict close-ups (to inspect and edit details) or overviews (to maintain "overview" of the entire document). Such views are usually shown on the same display simultaneously, e.g., in form of an overview inset that occludes small parts of a detail view. Navigation (zoom & pan) becomes significantly important in this scenario. It is usually accomplished directly within a view, e.g., by dragging the document with the mouse (pan) or by zooming in/out with keyboard shortcuts. With touch-enabled displays, a more direct way of interaction can be achieved. Yet, all interaction remains constrained to a 2D surface.

3.3 Our Goal: Bridging Some of the Gaps

In summary, the digital world offers a rich set of functionality with efficient ways to work with digital content. As a downside, it is restricted to a 2D workspace that provides only limited tangible affordances and thus does not leverage spatial arrangement as we are used to in the analog world. Aiming at closing this gap, we will next introduce the TUIP concept that combines some of the beneficial features of both worlds.

4 The TUIP Concept

The concept of Tangible User Interface Palettes (TUIP) is based on the traditional painter's metaphor, where a painter uses real-world tools like brushes and color palettes that are physically separated from the painting. With TUIP, we apply this idea to a digital tabletop. The tabletop screen shows graphics documents that users can edit by using digital pen input (see Figure 1). While tools (e.g., in form of menus) and the graphics document usually share the same screen space, we propose to decouple this space by making digital tool palettes physically tangible. We achieve this by using spatially aware handheld displays (Tangible UI Palettes) in different sizes and shapes that serve as physical representations for the otherwise virtual palettes. Users can take them into their hands and move and arrange them freely in 3D space and work with them using digital pen input. Thus, they bring back some of the advantages of a 3D work environment known from the analog world.

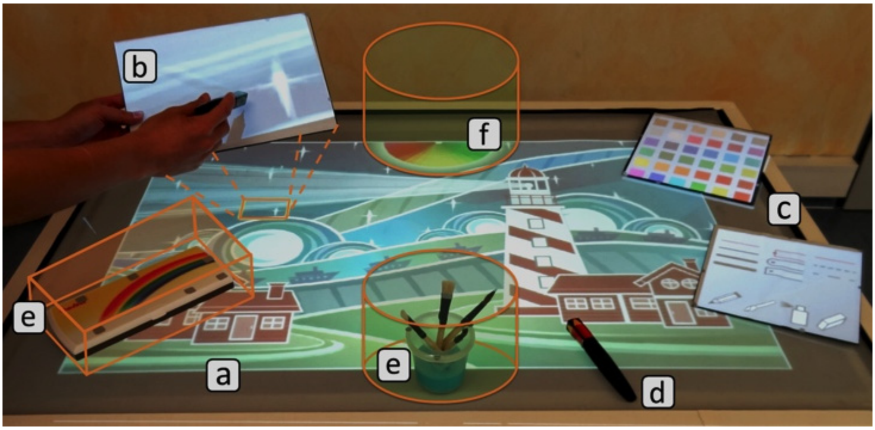


Fig. 1. Principle TUIP components: a tabletop with a graphics document (a), a handheld display with a detail view of that document (b), and handheld displays with tool menus (c). All displays are digital pen-enabled (d). Spatial work zones (e, f) are used as physical containers that can hold tools or document. They are represented by tangible proxies (e) or on-screen (f).

As all displays are dynamically updated depending on their position and orientation, instant visual feedback is displayed on them to reflect the state of tool usage. We take this concept further by introducing spatially bounded 3D interaction spaces (Work Zones) that exist above or beside the table. We use them as physical containers to group related functionality and documents. These work zones are visually represented either by tangible (physical) objects (see Figure 1e) or on-screen on the tabletop (see Figure 1f). Another important TUIP feature is digital pen input on all handheld displays and the tabletop, e.g., to simulate physical brush usage as known from the analog world. While in our work we only address workspaces with horizontal digital tables (standing and sitting usage), many of our concepts are general enough to be of use for other setups, e.g., tilted tabletops or wall-displays, as long as a horizontal surface for depositing handheld displays, pens and other physical objects is provided. In the following, we will discuss the TUIP design space and the concept of spatial Work Zones in more detail.

4.1 Design Space

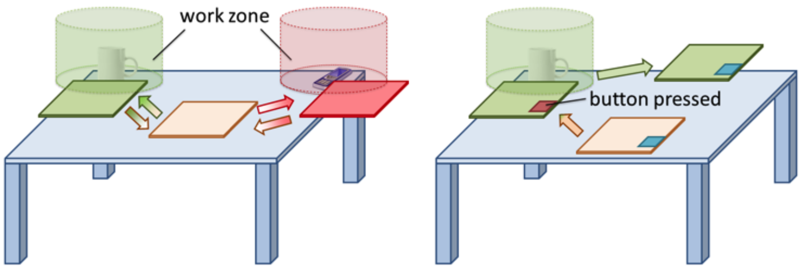
The TUIP design space is illustrated in Figure 1. The main components are a single tabletop (see Figure 1a) and multiple handheld paper-like projection screens (see Figure 1b,c) that are spatially tracked. These displays not only provide global and local views into a 2D information space, such as a graphics document (see Figure 1a,b), but can also switch dynamically to show functionality, e.g., in form of GUI palettes (see Figure 1c). For interacting with this system, we consider and support two input modalities: pen input and spatial input. **Pen input** is performed directly on the surface of tabletop and handheld displays with two degrees of freedom (2DOF). Technically, it is based on digital pen technology (Anoto), e.g., as it was proposed and used by Haller et al. for the Shared Design Space [13]. In our system, all displays are equipped with this technology (Figure 1d). **Spatial input** refers to the interaction with handheld displays by moving and rotating them through the physical space above or beside the table. With six degrees of freedom (6DOF), a rich set of interaction techniques becomes available, e.g., as categorized in [30]. One example for this is moving a display up and down for zooming. In order to make spatial interaction more explicit, we incorporated pressure-sensitive **on-palette buttons**, which further extends the interaction vocabulary. We also provide 6DOF spatial tracking for tangible proxies (used to represent spatial Work Zones).

One major advantage of such design space is the support of different input modalities that are close to what users are familiar with from everyday life. In particular, this includes drawing with (digital) pens and the spatial arrangement and organization of painting media and drawing tools (handheld displays). Clearly, both input strategies address a particular interaction goal better than the other one could possibly do. In this way, we are able to get a step closer to the overall goal of making the interaction more natural. The use of orthogonal input strategies also decreases the need of mode switches that are often perceived as being distractive. This helps take away mental load from users, thus potentially setting free intellectual capacity for more productivity and creativity. The combination of tabletop and multiple handheld displays enables the simultaneous use of shared and personal views (for both document and tools) and thus facilitates parallel work and co-located collaboration.

4.2 Spatial Work Zones

We propose to extend the interaction with spatially aware tangible displays by using proxemic interaction principles. We do this by introducing the concept of spatial

Work Zones that are distinct spatial 3D regions above and around the table with individual spatial positions and extensions. They define an independent interaction space that can be explored by moving handheld displays into and through them. This either temporarily or permanently changes what is displayed on a handheld display (see Figure 2). In this way, spatial work zones serve as physical containers that we use for pooling related functionality or digital content.



(a) Temporal Assignment (b) Permanent Assignment

Fig. 2. Spatial Work Zones are spatially bounded regions above and beside the tabletop. They provide fast access to tool palettes and documents. Moving a handheld display into them instantly changes what is shown on the display. Leaving the Work Zone immediately restores the previous content (a). Pressing a special on-display button makes the change permanent (b). This allows for taking along a particular tool palette when leaving the Work Zone.

The spatial extension of a work zone is defined by its physical shape and size within the working environment. For simplicity, we usually use spheres with a fixed diameter of about 30 cm. Each Work Zone has a center or spatial position that defines its location within the physical world. Users can freely manipulate this position and thus can arrange spatial Work Zones in their working environment as and when required. For this purpose, we introduce the concept of proxy representations. We support two types: tangible proxies and on-screen proxies that provide different levels of affordance. Users can seamlessly switch between these representations and use them, for example, to get an overview of available work zones or to move them to a new location. In the following, we will discuss their properties in more detail.

Tangible Proxies. Work Zones can be represented by real physical objects, which are optically tracked so that their spatial position is known to the system. These can be dedicated objects (e.g., a pot with brushes, see Figure 1e) only built and used for a specific purpose. They can also be generic objects like coffee mugs and post-its or other readily available objects. Tangible proxies offer a number of advantages. They are readily visible and thus can provide clues to the associated interaction. They are also freely moveable by physical means and allow for rearranging the associated Working Zone. This in particular includes the space beside the tabletop display, e.g., the table frame, where tangible proxies occlude no valuable screen space. Beyond that, personal objects like a wallet or mobile phone can be used to represent personalized Work Zones that provide, for example, access to confidential documents of a user or a customized tool configuration. One disadvantage of tangible proxies is that the system cannot actively move them, for example, to assist users by automatically restoring a previously defined Work Zone configuration. A current (technical) limitation is that visible markers need to be attached to each proxy object, though this could be solved in future iterations by relying on natural feature tracking techniques.

On-Screen Proxies. On-screen proxies are represented by regions situated on the tabletop display (see Figure 1f). Their advantage is that the system can change their location and appearance. This is useful, for example, when users want to restore their personal environment with a predefined Work Zone setup during the initialization phase or when the system needs to provide instant visual feedback about current state changes of the digital content associated with a Work Zone, e.g., the recently changed files of a Dropbox folder shared by multiple users. Users can freely arrange on-screen proxies on the tabletop display by performing pen- or finger-based drag gestures. As a downside, on-screen proxies usually occlude the document visible on the tabletop display. This problem can be lowered with well-established methods like partial transparency, context-sensitive fade-ins or showing them only when a handheld display moves towards them. Another problem is that spatial Work Zones represented by on-screen proxies are restricted to the direct proximity of the screen, thus limiting the user's freedom of arranging Work Zones farther away from the table. A possible approach to soften this problem is to use direction signs as a visual representation, e.g., an arrow could indicate that there is a Work Zone on the right side of the table.

Switching between Representations. A seamless switch between both types of representations can be achieved by performing a double-tap gesture with a tangible proxy on the tabletop screen. This means that the user must slightly knock on the tabletop surface with a physical proxy object in her hands. This works in both directions:

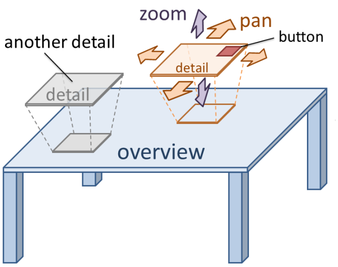
- **Screen-to-Tangible:** When the double-tap occurs directly on an on-screen proxy, its associated Work Zone will be transferred to the tangible proxy and the on-screen proxy is removed from the screen.
- **Tangible-to-Screen:** After a double-tap gesture is performed somewhere on the tabletop surface, an on-screen proxy is created at this position and the tangible proxy's Work Zone is transferred to this new on-screen proxy. This also implies that the tangible proxy is no longer linked to the Work Zone.

5 Use Cases

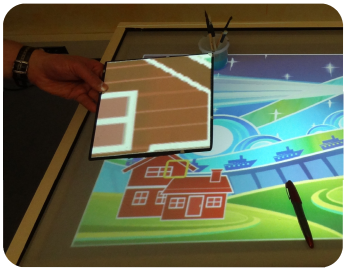
We continue by presenting several use cases that demonstrate how TUIP principles can support tool organization and document navigation using the example of graphics applications. Since TUIP combine techniques that have been presented previously, in particular Paper Windows [15] and PaperLens [29], we will first show how these fit into the overall concept (Use Case 1 & 2). We will then present what we consider our major contribution: the use of spatial Work Zones as a more transient way of working with tools and document views using handheld displays (Use Case 3 & 4).

5.1 Use Case 1: Document Navigation

To support the work with large graphics documents, efficient document navigation techniques are needed that allow users to efficiently zoom and pan (Challenge 2). We address this by physically detaching overview views from detail views and distribute them to the tabletop and several handheld displays (see Figure 3), similar as it was demonstrated by us in PaperLens [29]. In contrast to previous work, with TUIP we support both navigation (by moving a display) and drawing (by digital pen input). We therefore make use of explicit activation of spatial interaction, in our case, by pushing an on-display button. This guarantees that navigation only occurs when users really intend to, which is an important requirement for graphics document editing.



(a) Spatial input-based navigation (concept)



(b) Handheld zoomed-in view of a document

Fig. 3. Document navigation relies on spatial positions of handheld displays (up/down: zoom, horizontal: pan). This is similar to a concept proposed by us in PaperLens [29], except that we make the technique explicit (by pushing/holding an on-display button).

5.2 Use Case 2: Inter-display Transfer of Palettes

A major design goal of TUIP is to physically decouple menus from screen space in order to take them into the hands or to lay them out on the desk by still maintaining complete interactivity. This particularly addresses Challenge 1 (Tool Management). In this context, the seamless transfer of digital content between tabletop and handheld displays (in all thinkable combinations) is a crucial requirement. In Paper Windows [15], Holman et al. suggested to achieve this by placing a handheld display (in their

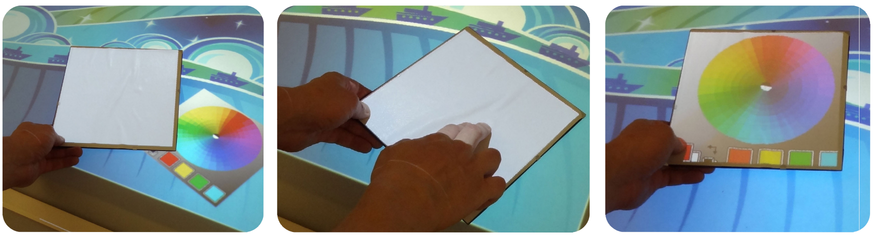


Fig. 4. As proposed by [15], the transfer of virtual windows from one screen, e.g., the tabletop (left), to another one (e.g., a handheld display) is done via a rubbing gesture (middle). This allows users to pick up digital tools (right) or parts of the document in a tangible manner.

case a digital piece of paper) directly on top of a virtual window and then to trigger the transfer by a rubbing gesture. We applied this concept to TUIP (see Figure 4).

5.3 Use Case 3: Quick Access to Tools by Using Work Zones

In the WIMP world, shortcuts such as keystroke combinations and mouse gestures provide quick access to specific interactions or frequently used tools. Inspired by the analog world, where artists collect their favorite pens and brushes in a coffee mug and put it right next to them on the desk for easy access, we propose to use spatial Work Zones (see Figure 2) for grouping related functionality (e.g., different color palettes for color management). Users can freely arrange these Work Zones on the table via tangible proxies, e.g., a jar with brushes (see Figure 1e).

A spatial Work Zone can contain multiple tool palettes simultaneously. We propose to organize them as a multi-layer stack (see Section 2.5). Moving a handheld display into a Work Zone triggers the multi-layer stack mechanism, i.e., no button is needed for any mode switch (see Figure 5a). The only temporarily visible tool palettes are instantly functional, e.g., ready to be used by a digital pen (see Figure 5b). As soon as the display is moved out of the Work Zone, the original content of the handheld display will be restored. This allows for making a series of adjustments in a streamlined manner, which can also involve more than one Work Zone, see Figure 5a.

(a) Switching between tool palettes by moving a handheld display from one Work Zone (left) into another one (right).

(b) Picking a color from a tool palette with a digital pen.

Fig. 5. By entering a spatial Work Zone with a handheld display the user gains quick access to available tool palettes (a). Any palette visible on a display is instantly interactive (b).

Permanent assignments (see Figure 2b) are achieved by double-clicking an on-display button when a desired GUI palette is visible. This allows users to put down a display on the table for other tasks, yet able to access the tool palette without going through the whole process again. A button press (and hold) prior to moving into a Work Zone achieves a transfer into the opposite direction. This lets users customize Work Zones by spatially "dragging-and-dropping" a GUI palette into the Work Space's multi-layer stack.

As proposed in [29,31], we visually guide users during the exploration of a multi-layer stack by fading in a height indicator on the right side of the display that shows all layers associated with a Work Zone (see Figure 6a). A red cursor bar in the height indicator shows the approximate position of the display within the stack. The height indicator automatically fades out when the user is not moving the display up/down for a certain time (2 seconds). We achieve a smoother user experience by blending adjacent tool palettes when a display reaches another layer of the stack (see Figure 6b).

(a) Height indicator (right) (b) Blending of adjacent layers

Fig. 6. A height indicator on the display provides visual orientation during the exploration of a multi-layer stack (a). The gradual visual blending of two adjacent palettes (when a display is about to cross a layer border) helps users to stay within a particular layer more easily (b).

5.4 Use Case 4: Quick Access to Graphics Layers by Using Work Zones

Spatial Work Zones can also provide quick access to other digital items commonly found in graphics editors, including parts of the graphics document. One example for this are graphics layers that store independent graphics components separately, thus allowing for a more flexible management of the document. The multi-layer stack matches this concept perfectly, e.g., by filling a dedicated Work Zone with all graphics layers of an image document. Similar to the previous use case, the exploration of individual layers is accomplished by entering the Work Zone with a display. Moving it up and down switches between layers. Here, different visibility configurations are possible, e.g., current layer only or all layers below the current one. When held vertically, the display provides fast access to an overview of all layers of a document, e.g., similar to Adobe Photoshop's Layer Panel. There is one problem, though: the work with vertically held handheld displays is not comfortable for prolonged usage. Fortunately, this can be tackled by making the overview permanent (e.g., by pressing an on-display button). This allows users to hold the display more conveniently, while they are working with it, e.g., to reorder or hide particular layers using pen or finger input.

By assigning a special tangible proxy to the "graphics layers" Work Zone, users can freely arrange the Work Zone to suit a particular task, e.g., by putting the proxy down somewhere on the table. This provides access to different image regions in a global vs. local scope. For example, when the Work Zone is positioned on top of a particular detail of the image (e.g., the left eye of a face), the Work Zone only provides access to layers involving that eye and its surroundings. In this way, users have quick access to relevant layers of an image detail. In contrast, when the "graphics layers" Work Zone is moved besides the tabletop display, the focus changes to global features, i.e., now scaled down miniature versions of layers are shown on the handheld display. This provides fast access to global layer management functionality, e.g., useful for organizing which layers are actually visible on the tabletop.

6 First Prototype

We have implemented a prototypic TUIP-system that relies on projective (paper-like) display technology, see Section 2.2. We use an infrared (IR)-based tracking approach, where all tracked devices (handheld displays and tangible proxies) have IR-reflective markers attached in distinguishable configurations. The system also includes a gesture recognizer to detect distinct move patterns (e.g., flipping gestures) of tracked devices. A ceiling-mounted projector is responsible for projecting digital image content onto handheld displays. Due to a skewed projection frustum, it also allows for projecting image content onto (almost) vertically positioned displays. A second projector projects from underneath the tabletop. An Anoto pattern is attached to both the tabletop and the handheld displays to support digital pens. Some of the handheld displays were equipped with a pressure-sensitive button (Arduino Xbee) that can wirelessly communicate state changes to the system. We use theses buttons primarily for explicit activation of spatial input-based navigation (see Figure 3a) and permanent assignment of tool palettes (see Figure 2b).

Considering the constraints of a research prototype, we did not spend much effort in implementing a real graphics editor application. Our focus was rather on illustrating the basic concepts of a TUIP system. This includes spatial Work Zones, spatial input-based navigation with handheld displays, and some basic tool palettes, e.g., for color and brush selection. Aside from this, pixel resolutions on handheld displays are rather limited due to the projector-based approach (about 65 pixels/inch). Thus, precision is by far not sufficient enough for serious work, e.g., as compared to modern tablets. However, the support of specific form factors of hand displays was more important to us. This includes small display weights (similar to cardboard), different display shapes and sizes, the avoidance of display bezels, the support of screens on the backside of displays, and the seamless support of pen-based drawing on all involved displays. Despite the technical limitations, we are confident that our prototype is adequate for basic testing of the underlying concepts. We expect that many of these limitations can be solved in the future, e.g., by using high-resolution projectors or organic light-emitting diode (OLED) display technology.

7 Initial User Experience

We have collected early user feedback on our prototype from four regular users of desktop graphics editors like Photoshop and Illustrator. All of them are experienced with touch- and pen-based interfaces. During an initial introduction to the prototype and its capabilities, we asked the participants not to get too much distracted by the rather low image quality or the limited functionality of the system. Rather, we asked them to assess the overall concept and whether they could image using similar techniques for regular work. All four users worked freely with the system for about 20 minutes. They performed a series of tasks requiring them to navigate within a graphics document and to switch between different tools frequently using TUIP techniques. After observing them interacting with the prototype, an interview was conducted to further elicit their experience and expectations for future use of the proposed system.

7.1 General Impression

Participants praised the seamless integration of pen-input on all displays as well as the support of spatial arrangement of digital content by using multiple handheld displays. Negative feedback was almost entirely related to the technical limitations of the prototype, but not to the underlying concept, which in general was assessed very positively. Although participants felt comfortable using the system for sketching and scribbling, they were (at this stage of implementation) reluctant in using it for more serious work, because (as they stated) this usually requires a much higher level of accuracy and a broader set of implemented tools, which was not yet supported by our prototype. However, users also said that they could imagine using a similar system for daily work once these problems are addressed properly. One user suggested that the system might be more preferable if the tabletop display could be tilted. Later, the same user acknowledged that a similar effect could be achieved by holding a handheld display accordingly and that maintaining a horizontal table surface provides adequate room for depositing pens, tangible proxies and handheld displays.

7.2 Document Navigation with Handheld Displays

Users particularly liked the possibility of decoupling specific regions of the image document from the main tabletop by transferring them to one or more handheld displays (see Figure 3). This allowed them to take a particular image detail into their hands, e.g., the left eye of a woman's face, while the context (e.g., the woman's face) was still visible on the tabletop. All participants stated that this considerably helped them to maintain focus and overview. Users also appreciated that image details on handheld displays were directly editable by pens, allowing them to draw in a zoomed-in view, while the overall image on the tabletop was updated live. Participants found it easy and even natural to select a specific image detail on a handheld display by moving the display through the air above the table (accompanied by pushing an on-display button to activate spatial input). After some minutes of practice, all users developed a fairly good understanding about the mapping of zoom level and height above the tablet surface. This effectively allowed them to directly jump to a specific zoom level by just holding the display at a particular height. Three users suggested that on-display buttons should provide tactile feedback to better reflect whether a button (and thus spatial input) was activated or not. Two users asked for additional support of touch-based navigation that they thought would be more appropriate for slightly re-centering a picture detail on a display when drawing with a pen.

7.3 Tool Organization with Spatial Work Zones

Participants described the concept of spatial Work Zones as intuitive and very useful for organizing digital content. In particular, they praised the ability to quickly switch between views of the document and tool palettes by simply entering/leaving one of the Work Zones. They also liked the possibility to freely arrange Work Zones on the table surface by using physical objects, including putting them away if not needed. One user suggested using Work Zones for representing special folders, e.g., as containers for "good" and "bad" photos in a photo collection. Users generally found it easy to pick one of the four tool palettes that we had attached to each Work Zone (organized as a multi-layer stack). However, users wished for more customization options regarding the number of palettes and their order within the palette-stack (something that we had not focused on in our prototype, yet). When working with the prototype, we observed that participants employed two distinct work patterns. The first pattern was employed when users were working on a single handheld display showing a detailed view of the document. Here, users approached spatial Work Zones repeatedly for quick (temporary) tools access (e.g., changing colors and pen thicknesses) by using temporary assignment (see Figure 2a). Participants stated that they particularly liked that the document view was instantly restored on the handheld display whenever they left a Work Zone. In the second pattern, users assigned two or three tool palettes to several handheld displays by using permanent assignment (see Figure 2b) prior to the actual drawing task. Then, they put the handheld displays on the table for convenient access. With that they used a digital pen to choose tools from handheld displays and to draw on the tabletop.

8 Conclusion and Future Work

In this work, we brought spatial affordances back to the digital world of graphics editing in a tabletop environment. Previous approaches have partially addressed this issue by augmenting digital tables with static (printed) physical palettes, for example, based on digital pen and paper technology. By using principles of spatial interaction and tangible affordances, we extended this idea by contributing dynamic (projected) Tangible UI Palettes and ways to access them. In particular, we proposed the concept of spatial Work Zones that are represented by tangible proxies or on-screen on the tabletop. We presented different use cases that illustrate the usefulness of our techniques. These use cases specifically addressed secondary interaction goals in graphics editing, like providing quick access to tool palettes and graphics layers, but can also be applied beyond this application domain. We implemented a prototype demonstrating our concepts. Early user feedback on the prototype was promising, especially concerning the ability to have fast access to digital content in a transient manner and the support for switching between tool menus and document views seamlessly on the very same handheld display.

In the future, we plan to drastically increase the set of available tools and enhance the visual quality and accuracy of the prototype with the goal of conducting a formal user study to further evaluate the efficacy of our techniques, including better support for inter-display interaction and more sophisticated techniques for spatial Work Zones. One particular advantage of the TUIP concept is that it is easily extendable and integrates well with multi-touch input, e.g., by using tablets. While in this work we primarily focused on spatial and pen input, we plan on integrating finger-based input, e.g., for precise zooming and panning of details on handheld displays.

Acknowledgements. This work was funded by the German Ministry of Education and Science (BMBF) project ViERforES-II (01IM1000 2B). We thank Hannes Witt for his help in implementing and evaluating the prototype.

References

1. Anoto Group AB – Website, `http://www.anoto.com`
2. Bailly, G., Lecolinet, E., Guiard, Y.: Finger-count & radial-stroke shortcuts: 2 techniques for augmenting linear menus on multi-touch surfaces. In: Proc. CHI 2010, pp. 591–594. ACM Press (2010)
3. Ballendat, T., Marquardt, N., Greenberg, S.: Proxemic interaction: designing for a proximity and orientation-aware environment. In: Proc. ITS, pp. 121–130. ACM Press (2010)
4. Bencina, R., Kaltenbrunner, M.: The design and evolution of fiducials for the reacTIVision system. In: Proc. Generative Systems in the Electronic Arts 2005 (2005)
5. Bier, E.A., Stone, M.C., Pier, K., Buxton, W., DeRose, T.D.: Toolglass and magic lenses: the see-through interface. In: Proc. SIGGRAPH 1993, pp. 73–80. ACM Press (1993)
6. Block, F., Haller, M., Gellersen, H., Gutwin, C., Billinghurst, M.: VoodooSketch: extending interactive surfaces with adaptable interface palettes. In: Proc. TCI 2008, pp. 55–58. ACM Press (2008)
7. Brandl, P., Leitner, J., Seifried, T., Haller, M., Doray, B., To, P.: Occlusion-aware menu design for digital tabletops. Ext. Abstracts CHI 2009, pp. 3223–3228. ACM Press (2009)
8. Dietz, P., Leigh, D.: DiamondTouch: a multi-user touch technology. In: Proc. UIST 2001, pp. 219–226. ACM Press (2001)
9. Elrod, S., Bruce, R., Gold, R., Goldberg, R., Halasz, F., Janssen, W., Lee, D., McCall, K., Pedersen, E., Pier, K., Tang, J., Welch, B.: Liveboard: a large interactive display supporting group meetings, presentations, and remote collaboration. In: Proc. CHI 1992, pp. 599–607. ACM Press (1992)
10. Greenberg, S., Fitchett, C.: Phidgets: easy development of physical interfaces through physical widgets. In: Proc. UIST 2001, pp. 209–218. ACM Press (2001)
11. Guimbretière, F., Nguyen, C.: Bimanual marking menu for near surface interactions. In: Proc. CHI 2012, pp. 825–828. ACM Press (2012)
12. Hall, E.T.: The Hidden Dimension. Anchor Books (1966)
13. Haller, M., Brandl, P., Leithinger, D., Leitner, J., Seifried, T., Billinghurst, M.: Shared design space: Sketching ideas using digital pens and a large augmented tabletop setup. In: Pan, Z., Cheok, D.A.D., Haller, M., Lau, R., Saito, H., Liang, R. (eds.) ICAT 2006. LNCS, vol. 4282, pp. 185–196. Springer, Heidelberg (2006)
14. Harrison, C., Dey, A.K.: Lean and zoom: proximity-aware user interface and document magnification. In: Proc. CHI 2008, pp. 507–510. ACM Press (2008)
15. Holman, D., Vertegaal, R., Altosaar, M., Troje, N., Johns, D.: Paper windows: interaction techniques for digital paper. In: Proc. CHI 2005, pp. 591–599. ACM Press (2005)

16. Holmquist, L.E., Mattern, F., Schiele, B., Alahuhta, P., Beigl, M., Gellersen, H.-W.: Smart-its friends: A technique for users to easily establish connections between smart artefacts. In: Abowd, G.D., Brumitt, B., Shafer, S. (eds.) UbiComp 2001. LNCS, vol. 2201, pp. 116–122. Springer, Heidelberg (2001)
17. Ishii, H., Ullmer, B.: Tangible bits: towards seamless interfaces between people, bits and atoms. In: Proc. CHI 1997, pp. 234–241. ACM Press (1997)
18. Izadi, S., Hodges, S., Taylor, S., Rosenfeld, D., Villar, N., Butler, A., Westhues, J.: Going beyond the display: a surface technology with an electronically switchable diffuser. In: Proc. UIST 2008, pp. 269–278. ACM Press (2008)
19. Johanson, B., Fox, A., Winograd, T.: The Interactive Workspaces Project: Experiences with Ubiquitous Computing Rooms. IEEE Pervasive Computing 1(2), 67–74 (2002)
20. Kakehi, Y., Naemura, T.: UlteriorScape: Interactive optical superimposition on a view-dependent tabletop display. In: Proc. TABLETOP 2008, pp. 189–192 (2008)
21. Kirsh, D.: The intelligent use of space. Artificial Intelligence - Special Volume on Computational Research on Interaction and Agency, Part 2 73(1-2), 31–68 (1995)
22. Kray, C., Rohs, M., Hook, J., Kratz, S.: Group coordination and negotiation through spatial proximity regions around mobile devices on augmented tabletops. In: Proc. TABLETOP 2008, pp. 1–8 (2008)
23. Kurtenbach, G., Buxton, W.: Issues in combining marking and direct manipulation techniques. In: Proc. UIST 1991, pp. 137–144. ACM (1991)
24. Lee, J.C., Hudson, S.E., Summet, J.W., Dietz, P.H.: Moveable interactive projected displays using projector based tracking. In: Proc. UIST 2005, pp. 63–72. ACM Press (2005)
25. Leitner, J., Powell, J., Brandl, P., Seifried, T., Haller, M., Dorray, B., To, P.: Flux: a tilting multi-touch and pen based surface. Ext. Abstracts CHI 2009, pp. 3211–3216. ACM Press (2009)
26. Morris, M.R., Ryall, K., Shen, C., Forlines, C., Vernier, F.: Beyond "social protocols": multi-user coordination policies for co-located groupware. In: Proc. CSCW 2004, pp. 262–265. ACM Press (2004)
27. Ryokai, K., Marti, S., Ishii, H.: I/O Brush: Drawing with Everyday Objects as Ink. In: Proc. CHI 2004. ACM Press (2004)
28. Song, H., Benko, H., Guimbretiere, F., Izadi, S., Cao, X., Hinckley, K.: Grips and Gestures on a Multi-Touch Pen. In: Proc. CHI, pp. 1323–1332. ACM Press (2011)
29. Spindler, M., Stellmach, S., Dachselt, R.: PaperLens: advanced magic lens interaction above the tabletop. In: Proc. ITS 2009, pp. 69–76. ACM Press (2009)
30. Spindler, M., Tominski, C., Schumann, H., Dachselt, R.: Tangible views for information visualization. In: Proc. ITS 2010, pp. 157–166. ACM Press (2010)
31. Spindler, M., Martsch, M., Dachselt, R.: Going Beyond the Surface: Studying Multi-Layer Interaction Above the Tabletop. In: Proc. CHI 2012. ACM Press (2012)
32. Sellen, A.J., Harper, R.H.R.: The Myth of the Paperless Office. The MIT Press (2001)
33. Ullmer, B., Ishii, H.: The metaDESK: Models and Prototypes for Tangible User Interfaces. In: Proc. UIST. ACM Press (1997)
34. Vandoren, P., Van Laerhoven, T., Claesen, L., Taelman, J., Raymaekers, C., Van Reeth, F.: IntuPaint: Bridging the gap between physical and digital painting. In: Proc. TABLETOP 2008, pp. 65–72 (2008)
35. Weiser, M.: The computer for the 21st century. Scientific American 265(3), 66–75 (1991)
36. Wilson, A., Benko, H.: Combining Multiple Depth Cameras and Projectors for Interactions On, Above, and Between Surfaces. In: Proc. UIST 2010, pp. 273–282. ACM Press (2010)

TIDE: Lightweight Device Composition for Enhancing Tabletop Environments with Smartphone Applications

Léo Sicard[1], Aurélien Tabard[2,1], Juan David Hincapié-Ramos[3,1], and Jakob E. Bardram[1]

[1] IT University of Copenhagen, Denmark
[2] University of Munich (LMU), Germany
[3] University of Manitoba, Canada
{leo.sicard,aurelientabard,jhincapie}@gmail.com, bardram@itu.dk

Abstract. Interactive surfaces like tabletop computers provide large touch-enabled displays, support novel forms of interaction and collaboration, and extend computation to new environments. However, being a novel platform, the existing application pool is limited and applications existing for other platforms have to be re-developed. At the same time, smartphones are pervasive computers that users carry around and with a large pool of applications. This paper presents TIDE, a *lightweight device composition* middleware to bring existing smartphone applications onto the tabletop. Through TIDE, applications running on the smartphone are displayed on the tabletop computer, and users can interact with them through the tabletop's interactive surface. TIDE contributes to the areas of device composition and tabletops by providing an OS-level middleware that is transparent to the smartphone applications, maintaining privacy by limiting content transfer between devices, and enhancing the usefulness of tabletops with already existing smartphone applications and software developers. We present the design and implementation of TIDE, the study of different interaction techniques to manipulate TIDE's interactive content, and an analysis of different research directions. Initial user feedback shows that TIDE is easy to use, learnable, and convenient for collaborative activities and private environments.

Keywords: Distributed User Interfaces, Multiple Display Environments, Tabletops, Smartphones, Device Composition.

1 Introduction

Tabletop computers have left the laboratories to become commercial products with rich input and output capabilities. Often used as appliances for specific purposes (e.g. exhibits, marketing events or demos), tabletops enable rich walk-up-and-use experiences. However, when building real-world applications for complex tasks [e.g. 6, 8, 29], designers must not only develop the application to support an activity (e.g. biology experiments, design meetings, etc.) but also support the set of basic expectations users would have from interacting with a computing device: the ability to browse the

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 177–194, 2013.

Web, access remote files, communicate, etc. The lack of applications to support such needs limits the impact and adoption of tabletops in professional environments.

At the same time, smartphones with rich computing capacities are now common, leading to a huge application pool and an important developer base. Particularly relevant to our work, is that many of the general-purpose applications already exist for the smartphone: Web browsing, communication and personal information management, access work related resources, etc. However, inherent to the mobile platforms are restrictions such as limited screen space, occlusion, and problems of touch-based interaction (e.g. the "fat finger" problem).

This paper presents *TIDE*[1] (*Tabletop Interactive Display Extension*): a novel approach to integrate smartphones to tabletops (see fig. 1). TIDE is a middleware running on both devices, allowing the smartphone to connect to the tabletop and replicate its user-interface onto the tabletop. The smartphone screen is displayed on the tabletop, and touch events on the tabletop surface are translated into touches on the phone display. We call this type of integration *lightweight device composition* because it is limited to display graphics and input mechanisms.

The main implication of this approach is that any smartphone application can be used on the tabletop. Moreover, smartphones and tabletops share a similar interaction model mostly based on touch and gestures, meaning that the applications do not have to be adapted to be used “through” the tabletop, or other bridging mechanisms be provided. At the same time, using the smartphone’s applications from the tabletop can mitigate some of the phone’s limitations, like hand obstruction or small form factor. Other implications are that the tabletop’s larger screen size and support for multiple users allow TIDE to better support tasks like reading, drawing or typing, or more social usages of smartphone applications.

Our contributions relate to technical, design, and usability aspects of TIDE. First, we present TIDE and its technical architecture. Smartphone applications run on the phone and are replicated on the tabletop; hence, there is no need for special application programming and any phone application can be replicated to the tabletop. Moreover, personal data never leaves the phone; hence increased security and privacy. Second, we present TIDE’s interaction design which is based on an elicitation study combined with our own design considerations, and present an evaluation of TIDE’s interaction techniques and their discoverability. Finally, we discuss the most convincing use cases based on an evaluation of our prototype, and lay out future research directions for TIDE.

Fig. 1. Tabletop users interacting with a map application running on a smartphone through TIDE

[1] Demo video: `http://www.youtube.com/watch?v=SAEARu-WRYk`

2 Background

Our motivation for designing TIDE comes from our experience building and deploying tabletops applications for professionals [28]. Such tabletop systems [6, 8] usually rely on a unique custom designed full-screen application. This means that most basic computing tasks are not supported or must be re-developed from scratch. To mitigate this problem the eLabBench [29] supports native Windows applications alongside its main application, while WeSpace [33] supports redirection of windows from laptops to shared displays. However, both the eLabBench and WeSpace only support native Windows applications and their WIMP style of interaction, which is particularly ill-suited to touch interaction.

This work is based on the observation that smartphones now offer a wide variety of touch-friendly applications for carrying everyday computing tasks. We thus developed TIDE to enable mobile applications to run on tabletops. In doing so, we push further existing concepts of smartphones' screen projection onto TVs or dedicated displays (e.g. [31]) through the notion of *lightweight device composition*. In this approach a host device (tabletop) allows a client device (smartphone) to use part of its screen to display content, and channel touch events to the client for processing. Based on our experience developing tabletop applications, lightweight device composition should support the following requirements:

R1. Enable applications to run without any modification.
R2. Support multiple client devices.
R3. Support walk-up-and-use scenarios with minimal set-up.
R4. Enable privacy control by letting users hide the client screen quickly.
R5. Support resizable applications (i.e. not only full-screen).
R6. Support physical separation between host and clients (e.g. putting the phone back in the pocket).

3 Related Work

As a middleware to integrate smartphones to tabletops, TIDE sits at the crossroad of research into the fields of device composition, tabletop augmentation and smartphone projection.

3.1 Device Composition

Inspired by the Ubicomp vision of seamless interaction between devices, device composition explores how heterogeneous devices can interoperate smoothly with each other. Initially under the name of "smart spaces", a number of projects (e.g. Augmented Surfaces [20], i-LAND [27] or Interactive Workspaces [9]) investigated this direction. From an architectural perspective, smart spaces were conceived as closed environments that rely on a centralized software infrastructure for coordination and control. Examples are BEACH, which supports the i-LAND project [30], and Gaia OS supporting Active Spaces [23]. The infrastructure facilitates the sharing of resources between the devices involved, including displays and peripherals, but also storage and computation. The advantage of working with a centralized infrastructure is that the interaction between devices can be optimized as they share a set of semantics (e.g. priorities, quality of service, etc.), and control mechanisms can take complex forms (authorization and authentication). However, the drawback is that new devices cannot be integrated easily, as they require specific software configurations.

More recent work on device composition aims at replacing centralized infrastructures by an ad-hoc peer-to-peer approach [11] or by creating a virtual or composite device. In the ad-hoc approach, devices communicate directly with each other to negotiate access to a certain resource (e.g. computer looking for a printer); each device makes decisions about resource allocation locally without the involvement of a central infrastructure. The virtual or composite device approach, seeks to aggregate resources from all the involved devices into a single virtual entity [2, 12, 13, 25]. Common to these two novel approaches is that they have new concerns aside from architecture of the underlying infrastructure, and include human-centered needs like user preferences [16] and manageability [17].

In TIDE, the smartphone and the tabletop come in direct communication with each other without depending on a central infrastructure, thus being closer to the second wave of device composition research. However, our approach differs from previous work in that it is lightweight: it is limited to the tabletop replicating the display of the smartphone and channeling touch events; neither the smartphone nor the tabletop can access any other resources, like computational power or storage, on the other device. Moreover, the composition is started only when the devices come into direct contact, i.e., the smartphone lies on the tabletop.

3.2 Tabletop Augmentation

The tabletop community has demonstrated repeatedly the value of going beyond simple touch interaction. For instance, augmenting tabletops with keyboards and mice can facilitate tasks for which touch input is not well suited like typing or selection of distant items [6]. Augmenting tabletops with tangibles enables users to control the state of tabletop applications with physical items and leverage the intrinsic properties of physical objects [10] or to display information attached to these objects [6]. Finally, tabletops have been integrated into larger ecologies of devices for tasks like meeting support [33], laboratory work [29] or collaborative search [14].

The most straightforward way to augment tabletop interaction with other devices is to leverage users' smartphones and transform them into tangible inputs. For example, Bluetable [34] connects wireless mobile devices to interactive surfaces using vision-based handshaking and Bluetooth. Phonetouch [24] combines image recognition with phones accelerometer data, by asking users to tap 3 times with their phone on the tabletop to initiate a Bluetooth connection between the two devices. Both projects allow for easy transfer of content (e.g. pictures) between the smartphone and the tabletop and from there to other smartphones.

TIDE differs from previous efforts to augment tabletops by focusing on giving access to the smartphone applications from the tabletop, rather than using the smartphone as a peripheral, as a control device or as a data source. Through the proposed lightweight device composition, TIDE not only provides the same functionalities of other smartphone-tabletop integration approaches (file sharing, picture viewing, etc), but also does so through the smartphone's interfaces the user is already familiar with.

3.3 Smartphone Projections

Another research line studied the opportunities of extending user-interfaces onto larger surfaces by projecting it either physically or virtually [19]. The recent availability of pico-projectors led to a number of projects projecting the screen of smartphones onto any surface. SixthSense [15] projects the interface on any object right in front of the user. Winkler et al. use the projected interface to provide extra features to the phone application [35], for example it extends phone calls with synchronous remote collaboration features on the projected interactive surface. Virolainen et al. enhance a docking station with a projector-based vertical FTRI touch-display [31] in order to increase the screen size of the phone. Nonetheless, pico-projectors have intrinsic properties which affect their usage: they need to be held in a specific and stable manner to provide a convenient interactive space, their luminosity and resolution is limited which inhibits rich UIs, and their energy consumption is significant.

Virtual projection is an alternative to circumvent some of the limits of pico-projectors. In its simplest form, a number of commercial smartphones can mirror their display on Television screens (e.g. iOS devices with AppleTV's AirPlay feature[2], Sony's Xperia devices[3], etc.). However this only projects the phones' output to the TVs and does not support input from the TV to the phone. Baur et al. proposed to virtually project the smartphone UI onto a larger one [3]. Here, optical projection serves as a metaphor for multi-device interaction: the user positions the smartphone toward the desired target surface, and the surface renders the device's interface taking into account its position and orientation in a way that resembles a physical projection. While this is an interesting direction, the users' engagement with the target surface is limited as they need to hold the smartphone, a requirement which in the long term could cause arm fatigue and thus reduced usage. Moreover, the user interacts on the phone and not on the projected surface, something which affects the system's usability in a collaborative situation.

TIDE is inspired by the simplest form of virtual projection, but adds the possibility to interact with the content from the target surface. Furthermore, the user does not need to hold the smartphone to control the projection, providing a stable replication, reducing the muscular strain and liberating both hands for interaction. Moreover, the luminosity is normally higher on a tabletop than it is with a pico-projector.

---

[2] https://www.apple.com/appletv/airplay/

[3] http://www.sonymobile.com/gb/xperia/

4 Tabletop Interactive Display Extension - TIDE

TIDE relies on principles of lightweight device composition: it merely pushes the display of the phone to the tabletop and pushes touch input from the tabletop to the smartphone. TIDE's approach to device composition does not involve other resources like storage, computation, or peripherals. This section presents a technical overview of TIDE focused on the requirements for lightweight device composition.

To initiate the screen replication phone and tabletop must be paired. We aimed at making this process transparent, thus users simply have to put their phone on the tabletop (fig. 2-A). Once TIDE detects a smartphone on the tabletop surface, it tries to connect to it and asks the user to confirm that s/he wants to establish the connection via a simple dialog on the tabletop (fig. 2-B). After the user validates the pairing, the display of the phone is replicated on the tabletop (fig. 3). At this point the phone can be moved to the side, and the user can manipulate and interact with the replicated smartphone on the tabletop (R6).

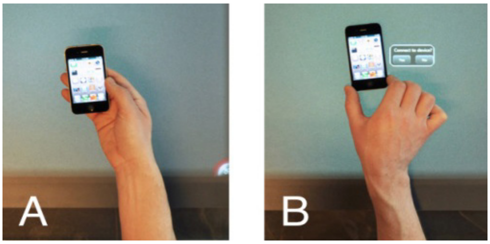


Fig. 2. Pairing a TIDE-enable iPhone to a Microsoft Surface tabletop

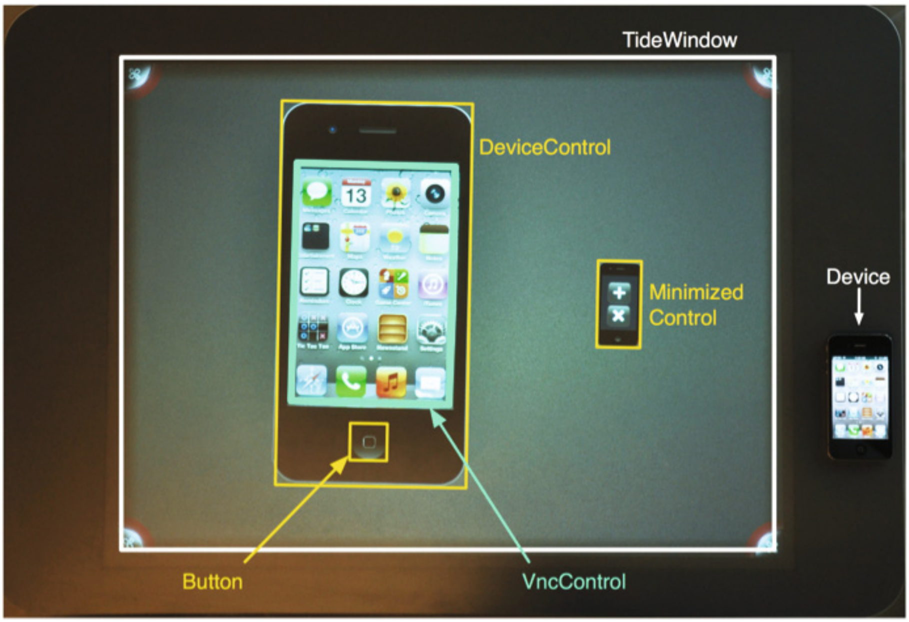


Fig. 3. TIDE Surface UI

4.1 System Components

At its core, TIDE relies on Virtual Network Computer (VNC) [21] to replicate the smartphone's user interface on the tabletop. A VNC server runs as a background process on the smartphone, and streams its screen-capture to a VNC client running on the tabletop (see green elements in fig. 4). Discovery and pairing relies on a vision-based tracking algorithm running on the tabletop (see blue elements in fig. 4). Finally, all user interactions are handled by standard WPF components for Microsoft Surface applications (see yellow elements in fig. 4).

Figure 3 shows the interactive elements of TIDE on the tabletop. The `TideWindow` is the main application window covering the whole surface of the tabletop. It contains UI elements that are enabled for touch-based interaction such as dragging, rotating and resizing. Each paired smartphone is associated with a virtual device implemented by a `DeviceControl` object which contains the replicated UI. The `DeviceControl` is responsible for detecting touch inputs that are destined to manipulate the replicated UI (i.e. capturing the touch events that are to be channeled to the smartphone). To provide consistency in the user experience, the `DeviceControl` has the visual aspect of the body of the smartphone. We considered less literal alternatives and discuss this choice in more details in section 5, but this one provide a set of benefits: users appreciated it and could quickly control the virtual phone e.g. scaling, rotating, hiding or accessing the physical buttons of the phone, such as the 'home' button of the iPhone.

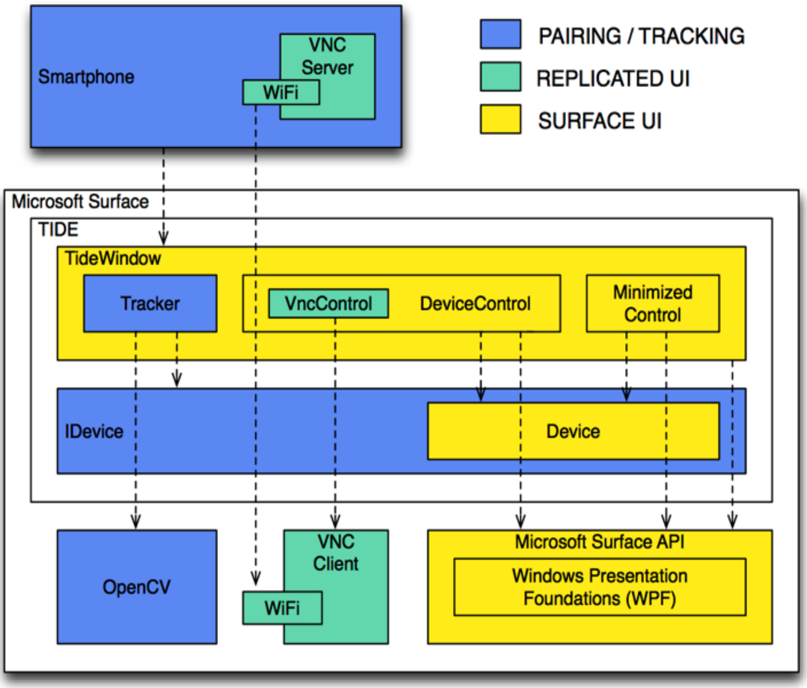


Fig. 4. TIDE component diagram

4.2 Implementation

Discovery and Pairing: The tracking component of TIDE is responsible for discovering smartphones by leveraging the MS Surface v1.0 infrared cameras. It uses shape recognition to discover new smartphones but also of keeping track of them during the application session. TIDE leverages the OpenCV shape recognition libraries to identify specific features of the phones' casings in order to track the phones' position and orientation on the tabletop. Once tracked, the pairing process relies on a simple approach which consists in attempting to connect to a series of known IP addresses on the local wireless network (R3). This simplistic approach enables an efficient discovery and pairing in testing and demonstration scenario, but would not scale well.

Replicated UI: We implemented UI replication with VNC in order to make TIDE compatible with all phone applications (R1). A VNC server runs on the smartphone, and the tabletop's TIDE implementation connects an embedded VNC client to it. The UI is streamed over Wifi, requiring the two devices to be connected to the same network. Once connected, the server sends a pixel-based replication of the phone's screen to the VNC client. TIDE receives the updates from the client via the VncSharp library and displays them on the tabletop. TIDE can create multiple client instances and connect to different smartphones (R2).

Surface UI: The Surface UI is implemented with the Surface SDK, which is itself based on WPF (Windows Presentation Foundations). The virtual phone is implemented by a `DeviceControl` object, which extends the Surface `ScatterView` control. The `ScatterView` class provides manipulation capabilities that are extensively used by the TIDE interaction model. The `DeviceControl` contains a `VncControl` object (from the VncLibrary), which displays the actual phone UI replication. Touch inputs to the virtual phone's screen are processed by the `VncControl` which maps their precise location on the screen and relays them to the VNC server on the phone.

4.3 Technical Limitations

Pairing and Discovery: TIDE uses shape recognition to discover new smartphones placed onto the tabletop. Once it detects a smartphone on the surface, TIDE tries to connect to a list of known IP-addresses for each device type (e.g. iPhone 4, HTC Desire, etc.) We took this approach, as the main focus of our work is the interaction with the replicated UI. However, a number of alternatives are available to provide a more robust discovery and pairing process in a real-life situation: Device discovery can be implemented using e.g. mDNS/Bonjour or UPnP and taking into consideration properties like phone type, location, and/or accelerometer data. Another approach would be to leverage the new Near Field Communication [32] of recent smartphones to exchange pairing information when being in close proximity of the tabletop.

UI Replication Protocol: TIDE builds upon VNC, which has the benefit of being available on most platforms and very stable. For instance, VNC is readily available on most mobile OSes, through third party applications. Moreover, because VNC is a pixel-based protocol is does not require any modification of existing applications.

However, VNC comes with a set of shortcomings, mostly due to the fact that it was designed to allow remote computing in a more "traditional" personal computer paradigm with keyboard, mice, WIMP applications, and stable/fast connectivity. Therefore, our implementation of TIDE only supports single-touch interaction. Moreover, VNC presents delays in the screen updates when a lot of visual changes take place on smartphones with high screen resolution (i.e. Retina displays).

Alternatives to mitigate these limitations include providing an improved VNC implementation and reducing features in the user-interface before registering small changes as updates.

Pixel Density: The resolution of modern smartphones is comparable to the ones of tabletops[4]. However the pixels per inch of the two classes of devices are significantly different: a tabletop pixel is considerably bigger than one from the smartphone. Moreover, when the replicated UI is bigger on the tabletop, the fidelity of the image is limited by the pixel density on the smartphone: a pixel on the phone is represented by several pixels on the tabletop; causing the replicated UI to look less sharp than it is.

Screen Size: Because of pixel-based replication and important differences in screen sizes, scaling up applications to take over the full screen of tabletops is not always valuable and depends on the type of applications. Applications based on default mobile widgets should be mostly used at a size equivalent to the one of the phone. Whereas applications with dense information, e.g. documents, maps or games can be run at larger size (e.g. full-screen) without suffering from the lower pixel density of tabletop screens.

5 Interaction Design Study

This section presents TIDE's interaction design, a process focused on discovering the main elements of the user experience: *how users would interact with the phone's representation on the tabletop*. Our goal was to maximize discoverability of the functionalities for novice users and foster spontaneous interaction. We took inspiration from Wobbrock et al.'s elicitation study of user-defined gestures for surface computing [29]. The authors conducted a study aiming at discovering the gestures (i.e. *actions*) participants tended to do to trigger specific behaviors (i.e. *commands*) on the tabletop, like scaling an image. Users were presented the *commands* and asked to perform the *action* they believe should be the cause.

Our approach is similar, but extends classical elicitation studies in that we wanted to incorporate the study results in a working prototype while maintaining a coherent set of commands. In our study, we thus took both a bottom-up approach (i.e. classical elicitation study, by asking participants what they would do to trigger a command) and a top-down approach in which we curated the actions before and after the study to maintain consistency among the different commands and actions.

---

[4] For instance, the resolution of Microsoft's PixelSense screen is 1920×1080, whereas recent Android phones often have a resolution of 1280×700 and the latest iPad a resolution of 2048×1536.

5.1 Interaction Design

Commands are the manipulations that can be completed on a TIDE window on the tabletop. We defined the following list of commands, as the minimal set needed for using TIDE:

1. *Dragging* the replicated UI across the interactive surface.
2. *Rotating* the replicated UI across the interactive surface.
3. *Resizing* the replicated UI across the interactive surface.
4. *Minimizing* the replicated UI, and restoring it.
5. *Hiding* the content of the replicated UI.
6. *Closing* the replicated UI (i.e. disconnecting.)

Actions are the way users can trigger the commands. Unlike Wobbrock et. al's study which focuses on gestures (i.e. no visual cues) [36], TIDE relies on visual elements. We created ***action controllers***, which are user-interface elements that allow the user to issue actions. We used sketches, storyboards and prototypes to envision different *action controller*. For each *action controller* there was a total of six possible *actions*. We focused on *action controller* that would be consistent with both the interaction experience on a smartphone and a tabletop (see fig. 5):

1. *Action Tabs* are traditional buttons/tabs that implement functionalities.
2. *Window Toggle* uses a switch to toggle the window between inactive and active states. In its inactive state the window can be handled like a digital picture.
3. The *Action Bar* is a manipulation area which resembles to a virtual touch-pad.
4. The *Active Border* is a digital frame around the window used for manipulation.
5. *Active Corners* is a strategy similar to Active Border, with the difference that the border's corners implement specific functionalities.

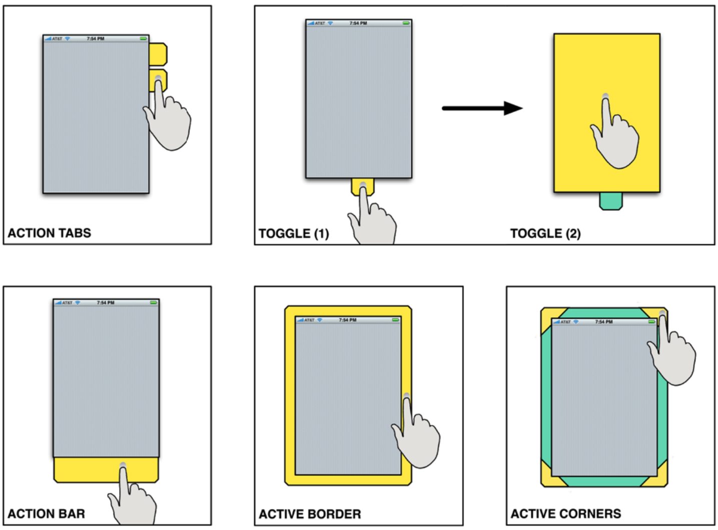


Fig. 5. Action controllers

For the purpose of the study, we created a 6th action controller called *'Other'* grouping *actions* that did not correspond to any of the five action controller, but were well suited to a specific *command*:

1. Drag by holding a finger on a specific tab, and using another finger to tap a destination target to move the window,
2. Rotate by performing a one finger dragging gesture on a corner of the window,
3. Resize by pulling the window apart with both hands,
4. Minimize by dragging the window to the bottom of the surface,
5. Hide by placing and holding a hand on the window,
6. Close by dragging the window to a specific location on the surface, and
7. Close by double tapping the window.

5.2 Elicitation Study

We conducted an elicitation study to investigate how participants would link the *actions* to the *commands*. We used paper mockups representing the phone interface, i.e., the replicated UI, and the action controllers (see [26] for details and images).

Participants: We recruited 12 participants aged between 25 and 35 years old. All of them were smartphone users. Three had prior tabletop experience.

Setting: An experimenter sat in a laboratory room with a participant, facing each other across a Microsoft Surface v1.0. A camera was used to record the sessions for future reference.

Procedure: Based on a script, the experimenter first presented an early version[5] of TIDE (non-interactive) to participants and the paper mockups of the iOS interface that participants could manipulate. The experimenter asked participants to envision writing an email on the tabletop with TIDE. This email-writing scenario was split into six steps. Each step required the user to issue a command. The experimenter used the following process to elicit every *action* from the participants:

1. Explain the desired command to the participant.
2. Ask first the participant to express an open-ended suggestion, i.e. suggest an action that s/he would perform to obtain the desired effect, and to demonstrate the action using the paper mockups.
3. Present three possible *actions* to issue the desired *command*, ask the participant to try them out to make sure they were understood. Then ask participants to rank the actions by order of preference.

Data Analysis: We ranked the {*command*, *action*} pairs based on the preference they received from participants. A weight of 3 was given to the first position, a weight of 1 to the second, and a weight of 0 to a third. We then aggregated the preferences of all the users and normalized them on an [0-1] interval, where a 1 meant that the entry was awarded a first position by all participants, and a 0 meant that all participants

5 Video of the early prototype:
`https://www.youtube.com/watch?v=EVWndZgTnPQ`

ranked the entry third. Table 1 summarizes the normalized scores, with colored cells containing values above 0.6, a score that can only be obtained if half the participants awarded it first position.

We also registered the participants' suggestions and counted how many users independently expressed a specific suggestion.

Table 1. Normalized weighted average rank given to each pair (command, action controller)

	Action Tabs	Window Toggle	Action Bar	Active Border	Active Corners	Other
Dragging	0.50	0.22	0.61	0.89	0.44	0.00
Rotating	0.06	0.17	0.56	0.61	0.78	0.50
Resizing	0.56	0.17	0.22	0.50	0.67	0.56
Minimizing	0.44	0.00	0.72	0.67	0.00	0.67
Hiding	0.33	0.39	0.17	0.78	0.44	0.56
Closing	0.28	0.50	0.06	0.50	0.33	1.00
Avg	0.36	0.24	0.39	0.66	0.44	0.55

5.3 Results

We can split the commands into two groups. Commands from the first group have a concrete visual signification, i.e., dragging, rotating and resizing. Commands from the second group are more abstract, i.e., minimizing, hiding, and closing.

For the first group, there is a strong coherence in the participants' choices. The favored *action controller* are the active border, the action bar and active corners. All three require the user to interact with an area directly around the window in order to manipulate it and modify its position, orientation or size. These interaction techniques are similar to the current standard for manipulating pictures on interactive touch screens. However in the present case, participants avoided touching the replicated UI because of its role as input relay between the tabletop and the smartphone.

For the second group, the action bar and active border also scored high, even though there is no apparent relation between the visual aspect of the strategy and the effect implied by the command. Looking closer at the results, participants preferred the double-tap for closing the window, possibly because it is a common technique in many other application contexts, as well as a quick and easy to execute.

The only '*Other*' strategies that scored above 0.6 are related to minimizing and closing the replicated UI. Both strategies, involve dragging the window to a specific location on the surface. This suggests that moving the window off screen is a natural way to remove focus from the application. Interestingly, this correlates with the analysis of the user suggestions, presented hereunder.

5.4 Participants' Suggestions

When asked to interact with TIDE for the first time, participants intuitively reached within the replicated UI to perform a dragging gesture with one or more fingers. However, any touch inside the replicated UI is forwarded to the smartphone. It was therefore necessary to stress again the distinction between replicated UI and surface

UI. Even though TIDE had been described to the participants their first reaction was still to actually interact with the phone's content. This finding suggests that for TIDE to be really intuitive, it should be able to interpret the intention of the user: to control the phone image or to control the phone remotely.

From the variety of ideas participants suggested, we identified a clear trend in two situations. For *resizing*, 8 out of 12 participants suggested grabbing the sides of the window with two fingers, and pulling the window apart to enlarge it. This shows that pinch and zoom is now completely part of users' vocabulary. For *minimizing*, 7 out of 12 participants suggested dragging the window off-screen (or to a specific location along the surface edge). The same suggestion reoccurred for the hiding and closing commands, although less strongly. This action is consistent with removing a real piece of paper from a table and appeared to be intuitive.

5.5 Final Design Choices

Based on the participants' rankings and suggestions we implemented the active borders and active corners (for manipulating and resizing – R5), and double tapping (for quick closing – R4). This selection represents the highest ranked *action controller* across all commands and the most popular suggested *actions*.

6 Usability Study

We conducted a final usability evaluation focusing on the learnability, ease of use and usefulness of the system. Our goal was to identify which *actions* and *commands* participants would discover in our working prototype

Participants: We recruited 10 participants aged between 25 and 40 years. All were regular smartphone users, and 5 had prior tabletop experience.

Apparatus: TIDE was installed on the Microsoft Surface v1.0, that has a 30 inch display (76cm) with a resolution of 1024 x 768 pixels. The tabletop is at the height of a coffee table, and the participants sat by it. It was used in combination with an iPhone 4 running iOS 5 and a HTC Legend running Android 2.1, both equipped with third-party VNC applications. An additional desktop computer was available for filling out a questionnaire.

Data Collected: In order to determine which *actions* were used to perform a given *command*, we used simple logging at the application level. Each session produced a set of CSV log files capturing the *command*, *action* and time. We also gathered participants' feedback through a questionnaire after each session, and the sessions were captured on video, for future reference.

6.1 Procedure

We first introduced the experiment to participants, explaining they would go through the following phases:

Exploration: The participants had three minutes to *learn by doing,* i.e. explore the system and discover its features on their own. We recorded which *actions* they discovered in order to evaluate system learnability for users with no prior knowledge of TIDE.

Guided Test: We asked participants to perform specific tasks with the application, in order to evaluate ease of use. The given instructions only included *commands*, and it was up to the participants to decide which *actions* to use.

Questionnaire: Participants filled out a questionnaire, allowing us to gather data assessing the usability and usefulness of TIDE.

6.2 Results

During the exploratory phase, all participants discovered at least one *action* for performing each *command*; suggesting TIDE is highly discoverable. Participants also felt subjectively that TIDE was easy to learn; this statement received a median score of 4.46 (sd=1.08) on a Likert scale from 5 Strongly Agree to 1 Strongly Disagree. The basic *actions* to *drag*, *rotate* and *resize* the application window were discovered by all participants in the exploration phase. This shows that tapping, dragging and pinching have now become part of a shared vocabulary among users.

When asked to select their favorite interaction techniques (*action* or *action controller*), the users did not necessarily choose the easiest one to discover. For instance, 8 users discovered the minimize *action* by resizing the window down, but only 1 selected it as a favorite way of activating this *command*. On the other hand, 6 users chose the double tap, which was as easily discovered, but quicker to perform. This finding emphasizes the importance of implementing multiple interaction techniques (*actions*) for a *command* as a good way to provide both discoverability and efficiency.

We evaluated the usefulness of TIDE by asking study participants, as well as 3 participants of the elicitation study, to assess in which situations and contexts they would

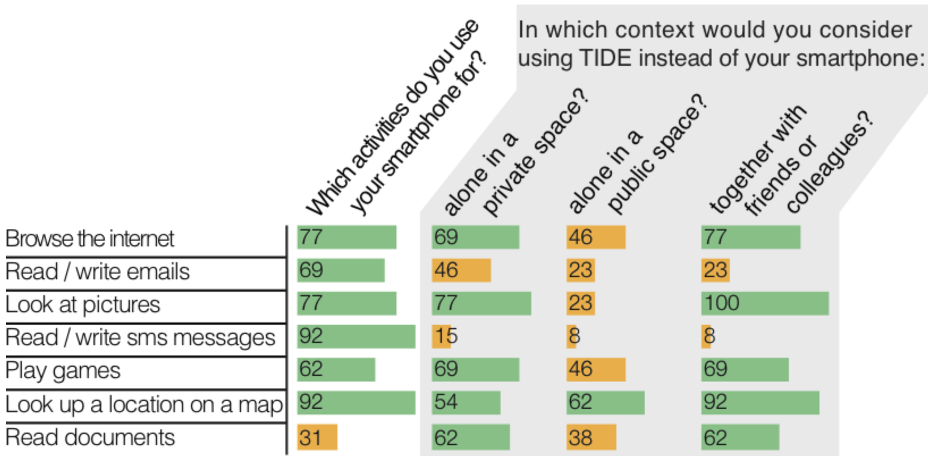


Fig. 6. The situations and contexts in which TIDE is found useful (Values expressed as percentage of the participants)

consider using the application. Figure 6 provides an overview of the situations in which TIDE made sense (or did not) to the participants. Participants' feedback clearly showed that a system like TIDE is not suited for interacting alone in a public space, nor does it seem well suited for interacting on tabletop like one would on a smartphone. The activities for which TIDE shows relevance are browsing the Internet, looking at pictures, playing games, looking at maps and reading documents; especially in collaborative situations. Such situations are not currently supported well by smartphones, not because of a lack of features but because of the inherent limitations of their small form factor.

7 Discussion and Perspectives

Participants' feedback outlines situations where TIDE would be useful in everyday life, like in private spaces or collaborative settings. This observation supports our initial motivation for developing TIDE as a platform to complement tabletop deployments in professional settings which are either private workspaces or semi-public ones with only known colleagues accessing the information. In this section we discuss the implications of TIDE for tabletop deployments in the real world, its benefits, and possible venues of future research.

7.1 TIDE in the World

Based on participants' informal feedback, we identified three situations beyond the professional environment in which participants would consider using TIDE:

1. At home, TIDE would play the role of a leisure device, to browse, play in a more public and shared manner than one would with a phone or a tablet.
2. In shared spaces, such as meeting rooms for collocated browsing of documents.
3. In public spaces but in a semi-public way, e.g. around a table with friends in a café or bar for casual browsing, playing, photos; while the space is public, the situation offers some degree of protection to intruders.

These situations add to our original motivation of providing touch-friendly applications in professional settings, to augment one's interactive desk.

7.2 TIDE Benefits

Compared to running tabletop based applications, by having applications run on users' smartphones, TIDE provides a set of interesting benefits:

1. **Users' personal data does not leave the phone**. This is of particular value given the shared nature of tabletops. By having applications run on the phone, users can be sure that their data is not shared without their consent, that it will not stay cached in the tabletop, and that it's not subject to eavesdropping during transport.
2. **No code coming from the tabletop runs on the users' smartphones**, by only transmitting input information from the tabletop to the phone and using a "*dumb*" protocol unaware of the content being transmitted, risks of accessing personal data are quite low.

3. TIDE provides users and developers with a **single-point of updates** to manage. Getting the latest version of an application happens only at the smartphone. This solves one of the usual problems of Ubicomp applications: the complexity of maintaining them and keeping them up-to-date.

7.3 Tabletop-Aware Phone Applications

Having applications installed on smartphones running through TIDE does not necessarily imply that such applications should be designed only for smartphones. We envision that TIDE could be extended so that applications running on the smartphone could be targeted for tabletop use. For example, UI elements could be adjusted to the number of pixels per inch of the tabletop, which is significantly lower than on smartphones. Other extensions could focus on the collaborative aspect of tabletops, for instance by allowing users to drag and drop data from on smartphone to another through TIDE.

This would mean extending VNC or developing a new framework dedicated to UI distribution like Substance [5]. Or like XICE [1], which is a programming framework supporting the development of applications enabling the annexation of displays by nomadic users.

7.4 Tabletop as a Hub for Peripherals

Finally we envision TIDE as a first step towards using tabletops as device composition hubs, offering extra computational power, faster connectivity, larger memory but also extending smartphones' input and output capabilities, by enabling users to plug-in external devices like keyboards and mice.

8 Conclusion

In this paper, we have introduced the notion of lightweight device composition and identified some of the requirements for its implementation. We presented TIDE, a middleware to provide such lightweight device composition between a smartphone and a tabletop computer. We presented the interaction design process of TIDE based on an elicitation study. We further carried out an evaluation to assess the discoverability of TIDE features. Finally we presented compelling cases in which TIDE could be used.

We observed that TIDE often generates a lot of questions about privacy related to the personal nature of smartphones. Nonetheless, after interacting with TIDE, users felt in control and did not voice any concern. This may be related to the fact that they had become familiar with the commands to hide or disconnect TIDE, and felt comfortable triggering them. In this sense, the privacy questions were rather at the application level, here users clearly stated that they would not use TIDE to view or write personal messages in a public setting, but would rather use TIDE for collaborative activities in which smartphones are not well suited. Moreover, we believe that social conventions play a role in controlling the privacy element, like it is the case today when a smartphone or tablet is shared among friends to show photos or play games.

References

1. Arthur, R., Olsen Jr., D.R.: Xice windowing toolkit: Seamless display annexation. ACM Trans. Comput.-Hum. Interact. 18, 14:1–14:46 (2011)
2. Bardram, J.E., Fuglsang, C., Pedersen, S.C.: Compute: a runtime infrastructure for device composition. In: Proc. AVI 2010 (2010)
3. Baur, D., Boring, S., Feiner, S.: Virtual projection: Exploring optical projection as a metaphor for multi-device interaction. In: Proc. CHI 2012. ACM (2012)
4. Chehimi, F., Rukzio, E.: Throw your photos: an intuitive approach for sharing between mobile phones and interactive tables. In: Proc. Ubicomp 2010, pp. 443–444. ACM (2010)
5. Gjerlufsen, T., Nylandsted Klokmose, C., Eagan, J., Pillias, C., Beaudouin-Lafon, M.: Shared substance: developing flexible multi-surface applications. In: Proc. of CHI 2011 (2011)
6. Hartmann, B., Morris, M.R., Benko, H., Wilson, A.: Augmenting Interactive Tables with Mice and Keyboards. In: Proc. of UIST 2009. ACM (2009)
7. Hartmann, B., Morris, M.R., Benko, H., Wilson, A.D.: Pictionaire: supporting collaborative design work by integrating physical and digital artifacts. In: Proc. of CSCW 2010 (2010)
8. Hunter, S., Maes, P., Scott, S., Kaufman, H.: MemTable: an integrated system for capture and recall of shared histories in group workspaces. In: Proc. of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2001, pp. 3305–3314. ACM (2001)
9. Johanson, B., Fox, A., Winograd, T.: The interactive workspaces project: experiences with ubiquitous computing rooms. Pervasive Computing 1(2), 67–74 (2002)
10. Jordà, S., Geiger, G., Alonso, M., Kaltenbrunner, M.: The reacTable: exploring the synergy between live music performance and tabletop tangible interfaces. In: Proc. of TEI 2007 (2007)
11. Karmouch, E., Nayak, A.: A distributed protocol for virtual device composition in mobile ad hoc networks. In: Proc. of ICC 2009, pp. 1–6. IEEE (2009)
12. Lee, J.W., Kim, S.M., Lim, H., Schuster, M., Domene, A.: A software architecture for virtual device composition and its applications. In: Ichikawa, H., Cho, W.-D., Satoh, I., Youn, H.Y. (eds.) UCS 2007. LNCS, vol. 4836, pp. 150–157. Springer, Heidelberg (2007)
13. Lyons, K., Pering, T., Rosario, B., Sud, S., Want, R.: Multi-display composition: Supporting display sharing for collocated mobile devices. In: Gross, T., Gulliksen, J., Kotzé, P., Oestreicher, L., Palanque, P., Prates, R.O., Winckler, M. (eds.) INTERACT 2009. LNCS, vol. 5726, pp. 758–771. Springer, Heidelberg (2009)
14. McGrath, W., Bowman, B., McCallum, D., Hincapié-Ramos, J.D., Elmqvist, N., Irani, P.: Branch-explore-merge: facilitating real-time revision control in collaborative visual exploration. In: Proc. of ITS 2012. ACM (2012)
15. Mistry, P., Maes, P.: SixthSense: a wearable gestural interface. In: ACM SIGGRAPH ASIA 2009 Sketches, p. 11. ACM (2009)
16. Mukhtar, H., Belaïd, D., Bernard, G.: User Preferences-Based Automatic Device Selection for Multimedia User Tasks in Pervasive Environments. In: Proc. of the 2009 Fifth International Conference on Networking and Services, ICNS 2009 (2009)
17. Newman, M.W., Elliott, A., Smith, T.F.: Providing an integrated user experience of networked media, devices, and services through end-user composition. In: Indulska, J., Patterson, D.J., Rodden, T., Ott, M. (eds.) PERVASIVE 2008. LNCS, vol. 5013, pp. 213–227. Springer, Heidelberg (2008)
18. Olwal, A., Wilson, A.D.: Surfacefusion: unobtrusive tracking of everyday objects in tangible user interfaces. In: Proc. of GI 2008, Toronto, pp. 235–242 (2008)

19. Pinhanez, C.: The everywhere displays projector: A device to create ubiquitous graphical interfaces. In: Abowd, G.D., Brumitt, B., Shafer, S. (eds.) UbiComp 2001. LNCS, vol. 2201, pp. 315–331. Springer, Heidelberg (2001)
20. Rekimoto, J., Saitoh, M.: Augmented surfaces: a spatially continuous work space for hybrid computing environments. In: Proc. of CHI 1999, pp. 378–385. ACM (1999)
21. Richardson, T., Stafford-Fraser, Q., Wood, K.R., Hopper, A.: Virtual network computing. Internet Computing 2, 33–38 (1998)
22. Rogers, Y., Hazlewood, W., Blevis, E., Lim, Y.-K.: Finger talk: collaborative decision making using talk and fingertip interaction around a tabletop display. In: Proc. of CHI E.A. 2004, pp. 1271–1274. ACM (2004)
23. Roman, M., Hess, C., Cerqueira, R., Ranganathan, A., Campbell, R.H., Nahrstedt, K.: A middleware infrastructure for active spaces. IEEE Pervasive Computing 1(4), 74–83 (2002)
24. Schmidt, D., Chehimi, F., Rukzio, E., Gellersen, H.: Phonetouch: a technique for direct phone interaction on surfaces. In: Proc. UIST 2010, pp. 13–16. ACM (2010)
25. Schuster, M., Domene, A., Vaidya, R., Arbanowski, S., Kim, S.M., Lee, J.W., Lim, H.: Virtual device composition. In: Proc. of the Eighth International Symposium on Autonomous Decentralized Systems, ISADS 2007, pp. 270–278. IEEE (2007)
26. Sicard, L.: TIDE - Using Device Composition on Tabletop Computers to Extend the Smartphone Experience, Master Thesis, IT University of Copenhagen (2012)
27. Streitz, N.A., Geissler, J., Holmer, T., Konomi, S., Müller-Tomfelde, C., Reischl, W., Rexroth, P., Seitz, P., Steinmetz, R.: I-land: an interactive landscape for creativity and innovation. In: Proc. of CHI 1999, pp. 120–127. ACM (1999)
28. Tabard, A., Hincapié-Ramos, J.D., Bardram, J.E.: The eLabBench in the wild: supporting exploration in a molecular biology lab. In: Proc. of CHI 2012, pp. 3051–3060. ACM (2012)
29. Tabard, A., Hincapié-Ramos, J.D., Esbensen, M., Bardram, J.E.: The elabbench: an interactive tabletop system for the biology laboratory. In: Proc. of ITS 2011, pp. 202–211. ACM (2011)
30. Tandler, P.: Software infrastructure for ubiquitous computing environments: Supporting synchronous collaboration with heterogeneous devices. In: Abowd, G.D., Brumitt, B., Shafer, S. (eds.) UbiComp 2001. LNCS, vol. 2201, pp. 96–115. Springer, Heidelberg (2001)
31. Virolainen, A., Paldanius, M., Lehtiö, A., Häkkilä, J.: Projector-based Multi-touch Screen for Situated Interaction with a Mobile Phone. In: Mobile and Personal Projection Workshop at CHI 2011 (2011)
32. Want, R.: Near Field Communication (**NFC**). IEEE Pervasive Computing, Smart Phone Dept. 10(3), 4–7 (2011)
33. Wigdor, D., Jiang, H., Forlines, C., Borkin, M., Shen, C.: WeSpace: the design development and deployment of a walk-up and share multi-surface visual collaboration system. In: Proc. of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2009, pp. 1237–1246. ACM (2009)
34. Wilson, A.D., Sarin, R.: Bluetable: connecting wireless mobile devices on interactive surfaces using vision-based handshaking. In: Proc. GI 2007, pp. 119–125. ACM (2007)
35. Winkler, C., Reinartz, C., Nowacka, D., Rukzio, E.: Interactive phone call: synchronous remote collaboration and projected interactive surfaces. In: Proc. of ITS 2011 (2011)
36. Wobbrock, J.O., Morris, M.R., Wilson, A.D.: User-defined gestures for surface computing. In: Proc. of CHI 2009, pp. 1083–1092. ACM (2009)

Evaluating the Effect of Phrase Set in Hindi Text Entry

Mohit Jain[1], Khushboo Tekchandani[2], and Khai N. Truong[3]

[1] IBM Research, Bangalore, India
mohitjain@in.ibm.com
[2] Dhirubhai Ambani Institute of ICT, Gandhinagar, India
[3] University of Toronto, Computer Science, ON, Canada

Abstract. Recently, many different Indic text entry mechanisms have been proposed and evaluated. Whereas the use of a common phrase set across text-entry research may help to produce generalizable results across studies, previous Indic Text entry evaluations have used a variety of different text entry phrases. In this paper, we develop and evaluate three different types of Hindi phrase sets that have been previously used in the literature – Hindi films, a grade VII textbook and a translated version of MacKenzie and Soukoreff's phrases – to study effects of their characteristics on performance. No statistical difference was found in novice user performance due to the different phrase sets. However, based on participant feedback, we report that consideration should be taken with regards to phrase length, frequency, understandability, and memorability in the design and selection of text-entry phrases.

Keywords: Hindi, Text Input, Phrase Set.

1 Introduction

Researchers are increasingly investigating how to best support text-entry for non-English languages, such as Indic languages [3], [5], [6], [7], [8], [9]. Indic languages such as Hindi, Marathi, Nepali, use Devanagari script. Hindi has 53 base letters – 34 consonants, 11 vowels and 8 diacritic marks. Unlike English, Hindi is not disconnected and mostly requires two or more letters to be combined together to form a *character* [6]. The combination of letters changes a character's visual appearance (*e.g.,* typing क(k) and ई(i) produces की(ki)).

The many challenges involved with entering text in Indic languages, including Hindi, have been discussed at length by others; for more detail, consult [5], [6], [7]. One of the key challenges is that for some characters, there are differences between the phonetic sequence of letters and the visual sequence of writing the letters (*e.g.*, character पि(pi) in the word पिता(pitaa) requires writing the vowel इ(i) before the consonant प(p), but pronounce the इ(i) after the प(p)) [6]. Furthermore, multiple base letters share a key button, often resulting in the use of multiple keystrokes to enter a

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 195–202, 2013.

letter (*e.g.,* typing Shift + L for थ). Many novel hardware and alphabetical layouts have been proposed to support effective Indic text entry on computers [3], [6], [8], [15] and mobile devices [5], [6], [9], [14], [17]. For evaluation, participants enter phrases of text using a technique while data is logged to calculate text entry performance. Due to the unavailability of a standard Hindi phrase set, researchers have used a variety of alternatives, including random words [9], random phrases [3], [6], paragraphs from a news portal [8], phrases from textbooks [5], and Hindi films dialogues [7]. Different phrase sets may have different properties, which can affect the text-entry task [21] and can potentially bias performance. Read [16] argues that when constructing text phrases, language and context must be taken into consideration.

In this research, we test three different Hindi phrase sets representing those that have been commonly used in previous studies. The phrases were taken from Hindi films (similar to [7]), grade VII Hindi textbook (similar to [5]), and a Hindi translated version of the MacKenzie and Soukoreff's phrase set [10] (state-of-the-art for English text entry) [http://www.dgp.toronto.edu/~mjain/HindiTextEntry.zip]. We show that the phrase sets differed in a variety of ways, such as the frequency of the words used, the length of the phrases, their correlation with a Hindi corpus, and the readability of the phrases. We conducted a study to examine whether these properties can bias performance of novice Hindi text entry users. We did not find any statistical difference exists in user performance, in terms of input speed and error rate, due to the different phrase sets. However, Mackenzie and Soukoreff [10] previously argue that use of a standard phrase set for English will help in attaining more generalizable results across studies. The same argument should hold true for a standard Hindi phrase set, and hence using these standard phrase sets for future Hindi text entry research will help in generalizing results. Additionally, based on participant feedback, consideration should be taken with regards to phrase length, frequency, understandability and memorability in the design of the phrases. In particular, our post-study questionnaire shows that ~8 words per phrase is an effective phrase length and familiar phrases minimize learnability and help motivate the participants.

2 Phrase Sets

We developed three different types of phrase sets previously used in the literature. Similar to [10], none of the phrase sets include any punctuation.

Phrases from Hindi Films (FP)

Previous Hindi text entry research has used phrases from Hindi films [7], as they are very memorable and capture the attention of the audience. We generated a phrase set comprising of 60 famous phrases from Hindi films. The phrases were randomly selected from online forums and blogs, citing famous Hindi film dialogues. Example phrases include: हम जहाँ खड़े हो जाते हैं लाइन वहीं से शुरू होती है (*where I stand, the queue begins from there*), डान को पकड़ पाना मुश्किल ही नहीं नामुमकिन है (*to catch Don is not only hard but impossible*).

Phrases from Textbooks (TP)

Grade VII textbook phrases have been used previously for evaluating Hindi text entry [5]. We generated a phrase set consisting of 50 phrases from a grade VII standard Hindi textbook [12]. We extracted phrases from a paragraph from 6 different stories in the book, thus maintaining some relationship between consecutive phrases. Example phrases include: दिव्या अनिल कि छोटी बहन है (*Divya is Anil's younger sister)*, यों तो वह शुरू से ही कमज़ोर है लेकिन इधर कुछ दिनों से उसे हर समय थकान महसूस होती रहती है (*we know she is weak from the start but the past few days she seems tired all the time*).

Translated MacKenzie and Soukoreff's Phrase Set (MSP)

The phrase set provided by MacKenzie and Soukoreff [10] is considered as standard and been used extensively for the evaluation of English-based text entry techniques. One of the authors, who is a native Hindi speaker, translated the phrase set into Hindi. In some instances, more context-appropriate words were used in place of English proper nouns; for example, *spaghetti*, *racketball*, and *Sam* were replaced with words with Indian context such as खिचड़ी (*khichdi* – an Indian food dish), क्रिकेट (*cricket*), and राम (*Ram* – an Indian name) respectively. Additionally, some English adjectives also do not have a direct translation in Hindi. We used the closest matching adjective with a high frequency in the Hindi language. Example phrases include: हाथी चूहों से डरते हैं (*elephants are afraid of mice*), प्यार के कई मतलब है (*love means many things*).

To verify the validity of the translated phrase set, two native Hindi speakers who are fluent in English were hired to rate the translation in three different parameters. The raters were provided with the English phrase set [10] of 500 phrases along with their Hindi translations. First, the raters were asked to rate the Hindi translated phrase according to this rating scheme – 1: *Directly Translatable* such that the English and the Hindi phrase mean exactly the same. 2: *Indirectly Translatable* such that synonyms must be used to keep the intended meaning (for example, "ginormous" to "very large"). 3: *Indirectly & Partially Translatable* such that words required changes to gain similar meaning (for example, "Sam" to "Ram", "spaghetti" to "khichdi"). 4: *Not Translatable* such that meaning is lost because the phrase contains idioms, colloquial expressions, etc. Second, these raters were asked to rate a Hindi phrase as "Y" if it is a valid Hindi phrase, "N" otherwise. A high-correlation was found between the two raters (95.45%). The rater rated 255 phrases as Directly Translatable, 173 as Indirectly Translatable, 12 as Indirectly Partial Translatable, and 60 as Not Translatable. Raters agreed that 444 phrases were valid Hindi phrases. For the study, we chose the first 150 valid phrases, which were rated as 1 by both the raters.

Linguistic Analysis of the Phrase Sets

An important characteristic that MacKenzie and Soukoreff [10] have established about their phrase set is its high frequency-based correlation with the English corpus. MacKenzie and Soukoreff's phrases vary from 16 to 43 characters (m=28.61). There are 2712 words (1163 unique) varying from 1 to 13 characters (m=4.46). It has a 0.945 *single-letter correlation* with the English letter frequencies reported by

Table 1. Linguistic Analysis of the Phrase Sets and Corpus

Metrics	EMILLE/CIIL Corpus	FP	TP	MSP	MS English Set
Number of phrases/sentences	737528	60	50	150	500
Number of words	12295677	490	673	881	2712
Number of unique words	202042	267	382	464	1163
Minimum word length	2	2	2	2	1
Maximum word length	33	10	13	14	13
Min. phrase length (# words)	1	4	3	3	3
Max. phrase length (# words)	888	14	39	11	9
Min. phrase length (# letters)	1	16	10	12	16
Max. phrase length (# letters)	4752	58	167	49	43
Single-letter correlation	-	0.97	0.98	0.98	0.95
Word-based correlation	-	0.70	0.68	0.75	0.85
Readability	m=10.34 sd=6.76	m=5.36 sd=2.4	m=8.0 sd=3.82	m=5.68 sd=2.46	m=4.17 sd=3.88
Words per phrase	m=16.67 sd=13.27	m=8.16 sd=2.4	m=13.46 sd=7.45	m=5.87 sd=1.6	m=5.4 sd=1.1
Letters per phrase	m=83.34 sd=67.4	m=35.45 sd=10.15	m=61.44 sd=34.63	m=26.82 sd=7.08	m=28.61 sd=5.02
Letters per word	m=4.06 sd=2.16	m=3.46 sd=1.44	m=3.63 sd=1.65	m=3.73 sd=1.72	m=4.46 sd=2.4

Mayzner and Tresselt [11]. *Word-based correlation* shows that 85.08% of the words in MacKenzie and Soukoreff's phrase set are from 1% of the total words from the COCA corpus (an English corpus that is recent to 2012: http://corpus.byu.edu/coca/).

The EMILLE/CIIL corpus [20] consists of 2387 Hindi files, totaling 737528 sentences and 12295677 words (202042 unique). Overall, all three phrase sets have a high word-based correlation with the EMILLE/CIIL corpus (0.70, 0.68 and 0.75 of the words in FP, TP and MSP, respectively, are from 1% of the total words from the corpus). This shows that the words used to build the phrase set are very frequently used words. The three phrase sets were similar in terms of single-letter correlation with the EMILLE/CIIL corpus (FP: 0.97, TP: 0.98, MSP: 0.98). Even the number of letters forming a word in each phrase set was not significantly different: FP: 3.46, TP: 3.63, MSP: 3.73. However, the phrase sets differed in terms of the number of words forming a phrase: FP: 8.16, TP: 13.46, MSP: 5.87 (Table 1).

An important characteristic of written text is readability, the ease with which text can be read. The lower the readability score of a phrase, the easier it is to read that phrase. The US Government Department of Defense uses Flesch-Kincaid Grade Level formula, which relies on the fact that the difficulty of a word is directly proportional to the number of syllables. Past Hindi research [1] has used this measure to evaluate readability. The readability attribute takes the effect of conjuncts into account.

To identify the number of syllables in a Hindi word, we used the syllabification algorithm proposed by [2] which has been reported to be more than 96% accurate. Our phrase sets readability level varies: FP: 5.36, TP: 8.0, MSP: 5.68. Phrases from grade VII textbook are found to be of grade 8 (sd=3.82), showing that the readability measure being used is efficient (Table 1).

Hypothesis

We hypothesize that (***H1***) use of MSP and FP will result in faster text entry and a lower error rate than TP. This is based on the fact that MSP and FP have lower readability, higher word correlation, and lower words per phrase (Table 1) compared to TP. Additionally, MSP's high word-based correlation to the EMILLE/CIIL corpus shows that it is very similar to what people read and write on a day-to-day basis; hence, we hypothesize that (***H2***) participants would prefer MSP over FP and TP.

3 Evaluation Method: Participants, Apparatus and Procedure

We recruited 18 individuals (mean age=21.8, 12 males, 4 females) to participate in a user study comparing the phrase sets. All were undergraduate Computer Science students, with an average 10.16 years of QWERTY keyboard usage. The criteria for selecting participants for the study was that they must know how to read, write, and speak in Hindi, but have never used an Inscript (Indian Script) keyboard [19] before. The mother tongue of all the participants, except two (who also know Gujarati), was Hindi. All participants, except one, received their primary education in English. Each participant was paid Rs 100 (~$2) per session. To motivate the participants, prize money of Rs 1000 and Rs 500 was given to the fastest and second fastest participants respectively.

We used the standard Inscript keyboard layout [19], attached to a 15.4 inches laptop screen. We developed custom software in C# which presented a test phrase at the top of the screen and asked the user to type the same phrase into a text box below it. The software logs all keystrokes for later analysis.

A within-subject multiple session study was conducted. Each participant took part in three 45-minute sessions. In each session, the participant was presented with phrases from one of the phrase set, and was asked to enter the visible text as quickly and as accurately as possible. A session consisted of two 20-minute typing blocks with a break of 3-5 minutes between the blocks. A minimum gap of 2 hours and a maximum gap of 2 days was maintained between two consecutive sessions. We counterbalanced the ordering of the phrase sets among the 18 participants, to allow learning to equally affect all conditions. Phrases within a set were always in the same order. After each session, participants were required to complete a questionnaire, rating the phrase set in terms of memorability, understandability, phrase length, and frequency of usage (phrase set measures were taken from previous research [10], [21]), on a 5-point Likert scale. For instance, for understandability 1 was very difficult to understand and 5 was very easy to understand.

4 Study Results

All participants completed the 3 session study. In total, together they entered 1340 phrases (388 FP, 310 TP and 642 MSP).

4.1 Speed and Accuracy

Text entry speed is measured in terms of words per minute (wpm), calculated using (letters per second)*60/5, with the definition that a word consists of 5 letters [18]. This definition was used to make the results more generalizable with the previous studies, and also to provide context for a larger audience to be able to relate the obtained results with English text entry speed; this has been done for Korean text input as well [4]. The average text entry rate over all the sessions across all the participants was 6.91 wpm (sd=2.43) (FP: m=6.22, sd=2.16; TP: m=7.28, sd=2.62; MSP: m=7.22, sd=2.48). Contrary to our hypothesis *H1*, no significant difference ($F_{2,34}$=2.5, p=0.1) was found between the three phrase sets.

Two metrics were used to measure text entry accuracy: (a) Keystrokes per Letter (KSPL) [18] – the number of keystrokes required to input a letter in Hindi, (b) Minimum String Distance (MSD) [18] between the presented and transcribed phrase. MSD accounts for the uncorrected errors in the final transcribed text; while KSPL measures the corrected errors as every correction adds multiple keystrokes, i.e., delete letter, re-enter letter. Note that for English, ideal KSPL value is 1 [18]. For Hindi, several letters require multiple keystrokes; we calculated and found the ideal KSPL for Inscript keyboard to be 1.12. For the entire study, a low average KSPL of 1.41 (sd=0.16) was found (FP: m=1.41, sd=0.13; TP: m=1.40, sd=0.1; MSP: m=1.43, sd=0.22). The average MSD value was 0.035 (sd=0.02) (FP: m=0.028, sd=0.01; TP: m=0.046, sd=0.03; MSP: m=0.03, sd=0.01). The higher MSD value for TP could be because the textbook phrases are not only longer than FP and MSP, but also have a higher readability value. However, no significant accuracy difference (KSPL: $F_{2,34}$=1.3, p=0.3, MSD: $F_{2,34}$=2.4, p=0.1) was found between the phrase sets, disproving our hypothesis *H1*.

4.2 Post-session Questionnaire

Because MSP preserves many of the positive qualities found in MacKenzie and Soukoreff's English phrase set (i.e., short, easy to understand and memorable phrases), participants preferred it the most (Friedman $\chi^2(2)$=14.7, p<0.01) (*H2*). The preference is derived cumulatively from post-questionnaire rating of phrases in terms of memorability, understandability, phrase length, and frequency of usage (FP: m=3.87, sd=0.1; TP: m=2.78, sd=0.1; MSP: m=4.41, sd=0.08). However, we did not find any significant difference between FP and MSP in terms of participants rating for understandability and memorability.

Understandability. Although participants perceived that the words used in the phrases were easy to understand across all the phrase sets, participants strongly agreed that FP and MSP contained phrases that are very easy to understand, while being neutral towards TP. Fisher's Exact test showed significant differences for understandability between FP (m=4.6, sd=0.8) and TP (m=3.1, sd=0.9) with $p<0.0001$, and between TP and MSP (m=4.17, sd=1) with $p=0.01$. This could be because TP has the smallest word-based correlation of the three phrase sets, but the highest readability level (where a high readability score means that the text is hard to read).

Length. Whereas participants agreed that phrases in TP were too long to type, they were neutral towards FP and somewhat disagreed that phrases from MSP were too long. Thus, around 8 words per phrase seems an acceptable length for phrases, but this might be influenced due to high understandability and memorability of FP.

We found a strong correlation between frequency of phrases and phrase length, with Pearson's $r(52) = 0.57$, $p<0.0001$. Participants thought that MSP contained very frequently used phrases but somewhat disagreed that FP and TP contained very frequently used phrases. Thus, seemingly longer phrases may not represent what participants would associate as frequently used phrases.

Memorability. Recently, Vertanen and Kristensson [21] argue that phrases should be memorable to help participants type them after only reading once. Participants strongly agreed that FP and MSP contained phrases that are very easy to memorize, while being neutral towards TP. We found that participants were able to quickly recognize the movies the phrases from FP were from and enjoyed typing them: *"Phrases should be interesting, so that we enjoy typing."* Many would often read the phrases aloud and laugh while typing. Fisher's Exact test showed significant differences for memorability between FP (m=4.2, sd=0.2) and TP (m=2.5, sd=0.1), and between TP and MSP (m=3.9, sd=0.2) with $p<0.0001$. Also a strong correlation was found between memorability and understandability, with Pearson's $r(52)=0.64$, $p<0.0001$.

5 Conclusion

In this paper, we develop and evaluate three types of phrase sets for Hindi text entry: phrases taken from Hindi films, a grade VII Hindi textbook, and a Hindi translated version of the MacKenzie and Soukoreff's phrase set. Although these phrase sets have different characteristics (Table 1), we did not find any text entry performance difference resulting amongst them. Our results suggest the use of different types of phrases in previous Hindi text entry studies potentially did not significantly affect performance results. However, it has yet to be shown that there is perhaps no significant difference between any sets of phrases. Thus, use of these common phrase sets can help in attaining more generalizable results across future research papers and hence facilitate comparison of different text entry techniques. Evaluating additional types of phrase sets (*e.g.*, [13]) and different characteristics of the phrase sets is important future work.

Additionally, we learned from participant feedback that MSP was preferred, and the acceptable Hindi phrase length should be ~8 words/phrase, but ~13 is too long. Readability contributed to the acceptability of FP, as does memorability. Familiar phrases help motivate the participants and make the whole typing task fun; this can be important in studies which ask participants to complete a large number of typing sessions. Thus, factors like readability, memorability, and phrase length, should all be taken into consideration together while choosing a phrase set.

References

1. Agnihotri, R.K., Khanna, A.L.: Evaluating the Readability of School Textbooks: An Indian Study. Journal of Reading 35, 282–288 (1991)
2. Choudhury, M., Basu, A.: A Rule Based Schwa Deletion Algorithm for Hindi. In: Proceedings of KBCS, pp. 343–353. Vikas Publications (2003)
3. Dey, P., et al.: Voice Key Board: Multimodal Indic Text Input. In: Proceedings of ICML-MLMI, pp. 313–318. ACM Press (2009)
4. Ilinkin, I., Kim, S.: Evaluation of Text Entry Methods for Korean Mobile Phones, a User Study. In: Proceedings of CHI, pp. 2023–2026. ACM Press (2010)
5. Joshi, A., et al.: Design and Evaluation of Devanagari Virtual Keyboards for Touch Screen Mobile Phones. In: Proceedings of MobileHCI, pp. 323–332. ACM Press (2011)
6. Joshi, A., et al.: Keylekh: A Keyboard for Text Entry in Indic Scripts. In: Extended Abstracts of CHI, pp. 928–942. ACM Press (2004)
7. Jung, Y., Joshi, D., Narayana-Saroja, V., Desai, D.P.: Solving the Great Indian Text Input Puzzle: Touch Screen-based Mobile Text Input Design. In: Proceedings of MobileHCI, pp. 313–322. ACM Press (2011)
8. Krishna, A.: Gesture Keyboard – User centered Design of a Unique Input Device for Indic Scripts. HCI International (2005)
9. Kumar, A., Shah, H., Joshi, A.: Saral: Devanagari Text Input System in Mobile Phones. In: Proceedings of UbiComp, pp. 1–4. ACM Press (2009)
10. MacKenzie, I.S., Soukoreff, R.W.: Phrase Sets for Evaluating Text Entry Techniques. In: Proceedings of CHI, pp. 754–755. ACM Press (2003)
11. Mayzner, M.S., Tresselt, M.E.: Table of single-letter and digram frequency counts for various word-length and letter-position combinations. Psychonomic Monograph Supplements 1, 13–32 (1965)
12. NCERT Books, `http://ncertbooks.prashanthellina.com`
13. Paek, T., Hsu, P.: Sampling representative phrase sets for text entry experiments: A procedure and public resource. In: Proceedings of CHI, pp. 2477–2480. ACM Press (2011)
14. Panini Keypad, `http://www.paninikeypad.com`
15. Rathod, A., Joshi, A.: A Dynamic Text Input scheme for phonetic scripts like Devanagari. In: Proceedings of Development by Design. ThinkCycle Press (2002)
16. Read, J.C.: Deceit, Distortion and Decision – Choosing Phrase Sets for Text Entry Research, `http://www.chici.org/references/deceit_distortion_and_decision.pdf`
17. Shanbhag, S., et al.: An Intelligent Multi-layered Input Scheme for Phonetic Scripts. In: Proceedings of SmartGraph, pp. 35–38. ACM Press (2002)
18. Soukoreff, R.W., MacKenzie, I.S.: Metrics for Text Entry Research: An Evaluation of MSD and KSPC, and a New Unified Error Metric. In: Proceedings of CHI, pp. 113–120. ACM Press (2003)
19. TDIL Inscript Layout, `http://tdil.mit.gov.in/pdf/keyboard/Keyboard%20Standard.pdf`
20. The EMILLE/CIIL Corpus, `http://catalog.elra.info/product_info.php?products_id=696`
21. Vertanen, K., Kristensson, P.O.: A Versatile Dataset for Text Entry Evaluations Based on Genuine Mobile Emails. In: Proceedings of MobileHCI, pp. 295–298. ACM Press (2011)

Frequent Words Improve Readability and Short Words Improve Understandability for People with Dyslexia

Luz Rello[1,2], Ricardo Baeza-Yates[1,4], Laura Dempere-Marco[3], and Horacio Saggion[2]

[1] Web Research Group,
[2] Natural Language Processing Group,
[3] Computational Neuroscience Group, Universitat Pompeu Fabra, Barcelona
[4] Yahoo! Labs, Barcelona
{luzrello,rbaeza}@acm.org,
{laura.dampere,horacio.saggion}@upf.edu

Abstract. Around 10% of the population has dyslexia, a reading disability that negatively affects a person's ability to read and comprehend texts. Previous work has studied how to optimize the text layout, but adapting the text content has not received that much attention. In this paper, we present an eye-tracking study that investigates if people with dyslexia would benefit from content simplification. In an experiment with 46 people, 23 with dyslexia and 23 as a control group, we compare texts where words were substituted by shorter/longer and more/less frequent synonyms. Using more frequent words caused the participants with dyslexia to read significantly faster, while the use of shorter words caused them to understand the text better. Amongst the control group, no significant effects were found. These results provide evidence that people with dyslexia may benefit from interactive tools that perform lexical simplification.

Keywords: Textual accessibility, dyslexia, eye-tracking, lexical simplification, readability, understandability, word frequency, word length.

1 Introduction

Dyslexia is a neurological reading disability that is characterized by difficulties with accurate and/or fluent word recognition and by poor spelling and decoding abilities [49]. Secondary consequences include problems in reading comprehension and reduced reading experience that can impede vocabulary growth and background knowledge [22]. Since a great amount of information is presented as text, this condition makes more difficult to people with dyslexia to use well standard information systems. According to the literature, people with dyslexia specifically encounter problems with less frequent words and long words [11, 20, 52].

The goal of this paper is to study to which extent word frequency and word length impacts text readability and understandability for native Spanish speakers with dyslexia. *Readability* refers to the legibility of a text, that is, the ease with which text can be read. *Understandability* refers to comprehensibility, that is, the ease with which text can be understood. Likewise to lexical simplification, which often involves

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 203–219, 2013.

replacing difficult words by their simpler synonyms [2], we replaced nouns by synonyms with diverse lengths and frequencies to study its effect on readability and understandability.

This research is motivated by (1) the social relevance of its results and, (2) the challenge of revealing whether people with dyslexia could benefit from lexical simplification tools, such as a browser plug-in which allows interactive substitution of complex words by simpler synonyms.

First, dyslexia is universal and frequent. Depending on the language, the estimation of the prevalence of dyslexia varies from 10-17.5% for the population in the U.S.A. [21] to 8.6-11% for the Spanish speaking population [7, 23, 36].

Dyslexia-related difficulties are also shared by other groups with special needs, such as people with low vision [16]. Moreover, symptoms of dyslexia are common to varying degrees among most people [12]. Another example of how research on dyslexia is useful for a broader public is the overlap of dyslexic-friendly recommendations with more general textual accessibility recommendations [29]. Hence, this investigation is extensible to general usability problems and to other target groups.

Second, given that dyslexia is a disability that affects language, we can assume that textual accessibility can be approached from two dimensions: *form*, i.e. customizing the text layout, and *content*, i.e. simplifying the text. Although there are applications that modify the text layout for users with dyslexia, such as *Claro ScreenRuler Suite* [8], *SeeWord* [17] or *IDEAL e-Book reader* [26], these only modify its design but not its content. Our experiments emulate possible lexical simplification strategies [2] and our results motivate the development of tools that simplify the textual content for people with dyslexia.

To the best of our knowledge, this is the first time that lexical complexity is measured in terms of readability and understandability for people with dyslexia using a methodology that combines eye-tracking and questionnaires. Therefore, this paper presents the following main contributions:

- A study about the effect of word frequency in readability and understandability, where the main result is that frequent words improve readability for people with dyslexia.
- A study about the effect of word length in readability and understandability, where the main result is that shorter words improve understandability for people with dyslexia.
- A comparison of both studies with a control group without dyslexia, where the effect of word length and frequency is not significant.

The rest of the paper is organized as follows. Section 2 covers related work while Section 3 presents the problem and the hypotheses of our research. Section 4 details the experimental methodology and Section 5 presents the results, which are subsequently discussed in Section 6. Finally, conclusions and future challenges are explained in Section 7.

2 Related Work

Since *readability* strongly affects text *comprehension*, sometimes both terms have been used interchangeably. However, previous research with people with dyslexia have shown that both concepts need to be taken into consideration separately. For instance, in [35] comprehension has been found to be independent of readability for people with dyslexia, while text readability can be used as an indicator of comprehension for people without dyslexia. Hence, in this study we distinguish between readability and understandability. Notice also that in the definition of dyslexia reading abilities and reading comprehension are treated separately [22].

Most research in this area focus on *English*, but our study considers *Spanish* language. Dyslexia manifestations vary depending on different language orthographies [5]. Since English and Spanish have different orthographies findings from one language do not necessarily apply to the other language. In fact, English has a deep orthography where the relationships between letters and sounds are inconsistent and many exceptions are permitted. On the other hand, Spanish has a shallow orthography with a more regular alphabetic system that contains consistent mappings between letters and sounds, the second shallowest of European languages [45].

We chose to study word frequency and word length because they are related to the word's processing time [33, 44], and they are strongly related to the difficulties that people with dyslexia find [11, 20, 52]. Previous work related to our research can be divided into: (a) studies from experimental psychology about eye-movement of readers with dyslexia, (b) work about comprehension abilities of people with dyslexia, and (c) interfaces for people with dyslexia.

Eye-Movements of Readers with Dyslexia. There is an extensive body of knowledge about eye-movements of readers with dyslexia using eye-tracking. Most of these studies focus on finding ways to diagnose dyslexia. However, the nature of eye-movements of people with dyslexia is currently under debate. While some found differences [14] among the two populations, others did not [47]. The closest work to ours comes from experimental psychology [20], which studies the effect of word length and word frequency in relation with eye fixation patterns among readers. Their results show that low frequency and long words present longer gaze durations and more re-inspections. Their approach focuses on finding eye-movement patterns to study particular words with the aim to discard the oculomotor dysfunction hypothesis of dyslexia [20]. They analyze single words while we analyze the whole text.

Comprehension in People with Dyslexia. Previous research has studied the effect of long sentences with difficult structures [46], sentence context [30], and the incorporation of graphical schemes [39], among others, on text comprehension of readers with dyslexia. More related to our work are [46, 43] who have suggested that the text could be made more difficult by the inclusion of low frequency and long words.

Interfaces for People with Dyslexia. There are number of existing applications for people with dyslexia that alter the presentation of the text such as *SeeWord*[1] [17],

[1] `http://www.computing.dundee.ac.uk/projects/seeword/`

Colour Explorer,[2] *Text4All*[3] [49], and *Penfriend XL.*[4] Other tools also incorporate text-to-speech technology such as *Claro Screen Ruler Suite,*[5] or the *IDEAL eBook reader*[6] [26]. *Dyseggxia*[7] [41] is a game for children with dyslexia that uses as a basis word modification, but, to the extent of our knowledge; there is no interface which modifies the text content to be read by people with dyslexia.

What is missing in previous research is knowledge about two factors of lexical complexity —word frequency and word length— in the Spanish language, and their impact on the reading performance and the comprehension of people with dyslexia, as well as the integration of these findings in tools for people with dyslexia.

3 Word Frequency and Length in Dyslexia

Among the language difficulties that people with dyslexia find, the additional difficulty that less frequent words *(e.g. pristine)* and longer words *(e.g. prestidigitation)* has been specially stressed [11, 20, 52]. The role of word frequency is so crucial in dyslexia that there is even a diagnosis of dyslexia based on the performance while reading frequent words [27].

Since word frequency and word length are naturally related in language, we studied them in two different experiments. The correlation originates from the fact that frequently used expressions tend to become shorter over time [24]. As stressed by Rayner *et al.* [33], to unveil cause and effect relationships, these dimensions have to be studied individually. Hence, our hypotheses are:

- **H1.1**: *A larger number of high frequency words increases readability for people with dyslexia.*
- **H1.2**: *A larger number of high frequency words increases understandability for people with dyslexia.*
- **H2.1**: *The presence of short words compared to long words increases readability for people with dyslexia.*
- **H2.2**: *The presence of short words compared to long words increases understandability for people with dyslexia.*

To the best of our knowledge, the rest of the dyslexia-related difficulties regarding words involve: irregular words[8] [9], orthographic similar words, *e.g. addition* and *audition* [15], homophonic words or pseudo homophonic words, *e.g. weather* and

[2] `http://colour-explorer.software.informer.com/9.0/`
[3] `http://www.text4all.net/`
[4] `http://www.penfriend.biz/pf-xl.html`
[5] `http://www.clarosoftware.com/index.php?cPath=348`
[6] `https://play.google.com/store/apps/details?id=org.easyaccess.epubreader`
[7] `https://itunes.apple.com/es/app/dyseggxia/id534986729?mt=8`
[8] Words in which there is no consistent correspondence between grapheme and phoneme, *e.g. vase*, pronounced as /vaz/.

whether [31], foreign words [11], new words, *e.g. fantabulous* [11], words with errors [9], non-words,[9] *e.g. happisfaction* [9], and numerical expressions [40]. In the experimental design we did not use words of such characteristics to control other possible effects.

4 Methodology

We study word frequency and word length as two independent variables in two different experiments. Nonetheless, the inherent relationship between frequency and length has constrained the selection criteria of the target words, i.e. long words are inevitably less frequent.

In the experiments, 46 participants (23 with dyslexia) had to read four texts, which were altered to include more/less frequent and longer/shorter words.

4.1 Design

For each experiment there were two conditions. In the first experiment, *word-frequency* [±frequent] served as independent variable with two levels: [+frequent] denotes the condition where suitable words were replaced by more frequent synonyms and [–frequent] denotes the condition where suitable words were replaced by less frequent synonyms. In the second experiment, *word-length* [±long] served as independent variable with two levels: [–long] denotes the condition where suitable words were replaced by shorter synonyms and [+long] denotes the condition where suitable words were replaced by longer synonyms.

The experiments followed a within-subjects design, so every participant contributed to each of the conditions in both experiments. The order of conditions was counter-balanced to cancel out sequence effects. To measure readability and understandability, we consider reading time, eye fixations duration, and a comprehension score as dependent variables. Next, we explain them in detail.

Reading Time. Shorter reading durations are preferred to longer ones since faster reading is related to more readable texts [50]. Therefore, we use *Reading Time*, i.e. the time it takes for a participant to completely read one text, as a measure of readability. This measure is extracted from the eye-tracking data.

Fixation Duration. When reading a text, the eye does not move contiguously over text, but alternates saccades and visual fixations, that is, jumps in short steps and rests on parts of the text. *Fixation duration* denotes how long the eye rests still on a single place of the text. Fixation duration has been shown to be a valid indicator of readability. According to [25,33, 43], shorter fixations are associated with better readability while longer fixations can indicate that processing loads are greater.

Hence, we use fixation duration as a readability measure in addition to the reading time.

[9] A non-word is a word that has no meaning, is not known to exist, or is disapproved.

Comprehension Score. To measure text comprehension we used inferential items, that is, questions that require a deep understanding of the content of the text. We used multiple-choice questions with three possible choices, one correct choice, one partially correct choice, and one wrong choice. To compute the text comprehension score, the choices counted 100%, 50%, and 0%, respectively.

4.2 Design

We had 23 Spanish speakers (12 female, 11 male) with a confirmed diagnosis of dyslexia (group D) that participated in our study. They were asked to bring their diagnoses to the experiment, to guarantee that dyslexia was diagnosed in an authorised centre or hospital. Their ages ranged from 13 to 37, with a mean age of 20.74 years (s = 8.18).

Three of these participants were also diagnosed with attention deficit disorder. All participants were frequent readers; per day, eleven read less than four hours per day, nine read between four and eight hours per day, and three participants read more than eight hours.[10] Ten people were studying or already finished university degrees, eleven were attending school or high school, and two had no higher education.

A control group of 23 Spanish speakers (13 female, 10 male) without dyslexia also participated in the study (group N). Their ages ranged from 13 to 35, with a mean age of 20.74 years (s = 8.18). That is, overall, we had 46 participants (25 female, 21 male).

4.3 Materials

To study the effects of word length and frequency, we need to study target words in context, that is, as part of a text. The rationale behind this is that readability and understandability pertain to longer segments of texts [19]. To isolate the effects of these variables, the texts need to be comparable in complexity. In this section, we describe how we designed the texts, the target words, and the questionnaires that were used in this study.

Base Texts. As basis for our tests, we picked four short texts with an average length of 60.63 words. For *word-frequency*, we used two texts about the consequences of wars in Pakistan and Somalia. These were the most similar pairs of texts we could find in the Spanish Simplex corpus [3]. For *word-length*, we created two mystery stories, one about a wizard and one about a car. In the following, we denote these texts with *Pakistan*, *Somalia*, *Wizard*, and *Car*.

To meet the comparability requirements among the texts belonging to the same experiment, we adapted the texts maintaining as much as possible the original text. We matched the readability of the texts by making sure that the parameters commonly used to compute readability [13], had the same or similar values:

---

[10] We think that the participants interpreted reading time as schooling time and that is why the numbers may seem exaggerated.

(a) They have the same number of target words: Fifteen [±frequent] word pairs for each of the texts in the *word-frequency* experiment and six [±long] word pairs for each of the texts in the *word-length* experiment. Only in *Somalia* (*word-frequency* experiment) two target words are repeated in the text.
(b) Within each experiment, the texts use the same genre, international news.
(c) They are about similar topics.
(d) They contain the same number of sentences: four sentences for each of the texts in the *word-frequency* experiment and three in the *word-length* experiment.
(e) The texts of each experiment have the same number of words per text (93 words for both texts in the *word-frequency* experiment and 33 words in the *word-length* experiment).
(f) All the texts have a similar word length, with an average length ranging from 4.89 to 5.50 letters.
(g) They contain the same number of unique named entities and they do not contain foreign words, numerical expressions or acronyms.

Text Layout. The presentation of the text has an effect on reading speed of people with dyslexia [18]. Therefore, we used a text layout that follows the recommendations of previous research. As font type, we chose *Arial, sans serif*, as recommended by Al Wabil *et al.* [1]. The text was left justified, as recommended by the British Association of Dyslexia [4]. Each line did not exceeded 62 characters/column, the font size was 20 point, and the colors used were black font with crème background,[11] as recommended by Rello *et al.* [42].

Target Words. Word length and frequency were selected taking into consideration the problems that people with dyslexia encounter. To control that other dyslexic-related difficulties did not interfere in the selected target words, we also took into account linguistic criteria during the selection of the target word pairs of synonyms:

(a) We did not include ambiguous names because they require more processing than unambiguous words [33].
(b) We did not change compound nouns or collocations in the texts. Instances like *secretario general, `secretary of state',* which is composed by two tokens but one meaning, were discarded.
(c) We did not use irregular words, foreign words, non-words, and new words [9, 11].
(d) We took into consideration only common names. Uncommon names were discarded because they are more likely to be irregular, foreign, or new words.
(e) We did not take into consideration phonetically similar words, that is, homophonic words or pseudo homophonic words [31] and orthographically similar words [15].

To apply our criteria (a) to (d) we used linguistic knowledge using the Royal Spanish Academy Dictionary [34]. To control for orthographic and phonetic similarity (e)

[11] The CYMK are crème (FAFAC8) and black (000000). Color difference: 700, Brightness difference: 244 [42].

of the target word, we consulted the database of indexes of frequency, length, and orthographic neighbours[12] in Spanish [32], which enables us to determine how many neighbours the target word has and how frequent are those neighbours. We can assume that orthographic and phonetic similarities are related in Spanish, because, as we mentioned before, it has a shallow orthography.

Word Frequency. For creating the pair of [±frequent] synonyms we first extracted the nouns of the texts according to the previous selection criteria. Then, we checked their synonyms using a synonym dictionary[13] and created a list of synonyms for each target word. We discarded nouns with no synonyms such as *million* or *kilometer*. We subsequently computed the relative frequencies of each of the synonyms for each of the lists using the advanced search of a major search engine.[14] Then, we manually selected the pair of synonyms (most frequent and less frequent) for the context where the target noun occurred in the text. Each pair of [±frequent] has a frequency difference of at least one order of magnitude, for instance *morada, 'house'*, is 67 times less frequent than *casa, 'house'*. In the Appendix we present the pairs of synonyms with their frequency ratio. To maintain both factors independent, short nouns with less than four letters were discarded. Still, we could only partially control for length differences: the average length per word in the high frequency synonyms is 7.62 letters while for the low frequency synonyms is 9.56 letters.

Word Length. Finding synonym pairs with a relatively large difference in word length is challenging for Spanish because 79% of the words in the dictionary have between 6 and 11 letters. Note that the average word length in Spanish is 8.78 letters per lemma, being similar to English where is 8.99 letters.[15]

Therefore, for creating the pair of [±long] synonyms we took all the Spanish lemmas from the Royal Spanish Academy Dictionary [34] and selected the longest words. Then, we looked up for the synonyms[16] of these words to create the lists of synonyms. However, the majority of the longest words in Spanish do not have synonyms, for example *electroencefalograma, 'electroencephalogra'*. From the list of synonyms we selected synonym pairs where the [+long] synonym at least doubled the length of its [–long] counterpart. For instance apartamento, *'flat'*, is 2.75 times longer than *piso, 'flat'*. In the Appendix, we present the pairs of synonyms with their length ratio.

Comprehension Questionnaires. We used multiple-choice questions with three possible choices, one correct choice, one partially correct choice, and one wrong

---

[12] Neighbours are all the words with the same length as the target word that differs in a single letter [10], that is, the Hamming distance between them is one. For instance, the word *casa* has many neighbours like *masa, cosa, cama* and *caso*.

[13] `http://www.wordreference.com/sinonimos/`

[14] `http://www.google.com/advanced_search`

[15] We used the Royal Spanish Academy Dictionary [34] and the Longman Dictionary of Contemporary English [28].

[16] `http://www.wordreference.com/sinonimos/`

choice. The questions were inferential, that is, answering it correctly required a deep understanding of the text content. An example question is given below:

(a) ¿De qué trata el texto? *'What is the text about?'*.

a. Sobre la acción de la Cruz Roja en Somalia.
   'About the Red Cross action in Somalia.'
b. Sobre las consecuencias de la guerra en la población de Somalia.
   'About the consequences of the war on the Somalia population.'
c. Sobre el incremento del precio de los productos en Somalia.
   'About the price increase of products in Somalia.'

4.4 Equipment

The eye tracker used was the Tobii T50 [48] that has a 17-inch TFT monitor with a resolution of 1024x768 pixels. The time measurements of the eye-tracker have a precision of 0.02 seconds. The eye tracker was calibrated for each participant and the light focus was always in the same position. The distance between the participant and the eye tracker was constant (approximately 60 cm. or 24 in.) and controlled by using a fixed chair.

4.5 Procedure

The sessions were conducted at Universitat Pompeu Fabra and lasted from 30 to 40 minutes, depending on the amount of feedback given by the participant. In each session, the participant was alone with the interviewer in the quiet room prepared for the study, and performed the following three steps.

First, we began with a questionnaire that was designed to collect demographic information. Second, we conducted the two experiments, which were recorded using eye-tracking. Then, the participants were asked to read the texts in silence and to complete the comprehension tests after each completed text. Finally, we carried out a semi-structured interview, where we asked for feedback about the experiment and about the usefulness of lexical simplification.

5 Results

In this section we present the analyses of the data from the eye tracker (reading time and fixation duration) and the comprehension tests. First, we analyzed the differences among groups (D and N) and then the effect of the conditions within each group.

The results are shown in Tables 1 and 2. A Shapiro-Wilk test showed that the three data sets were not normally distributed. However, a Barlett's test showed that they were homogeneous. Hence, we use Student's t-tests to show statistically significant effects.

Table 1. Results of the *word-frequency* experiment

Dependent Variable (ave. ± std.dev.)	[+Frequent]	[−Frequent]
	Group D	
Reading Time (s)	41.99 ± 13.00	53.35 ± 19.10
Fixation Duration (s)	0.22 ± 0.04	0.25 ± 0.05
Comprehension Score	90%	87.5%
	Group N	
Reading Time (s)	24.17 ± 6.18	29.25 ± 10.77
Fixation Duration (s)	0.17 ± 0.03	0.18 ± 0.03
Comprehension Score	95%	87.5%

Table 2. Results of the *word-length* experiment

Dependent Variable (ave. ± std.dev.)	[+Long]	[−Long]
	Group D	
Reading Time (s)	21.23 ± 10.28	13.74 ± 6.05
Fixation Duration (s)	0.23 ± 0.05	0.22 ± 0.04
Comprehension Score	88.64%	65.91%
	Group N	
Reading Time (s)	11.78 ± 4.24	9.59 ± 3.11
Fixation Duration (s)	0.18 ± 0.03	0.17 ± 0.03
Comprehension Score	90.91%	86.36%

Reading Time. We found a significant difference between the groups regarding reading time ($t(124.707) = 5.434$, $p < 0.001$). Group D had significantly longer reading times ($\mu = 31.86$, $s = 20.28$ seconds) than the participants of group N ($\mu = 18.32$, s = 10.53 seconds).

Word Frequency. For the participants with dyslexia we found a significant effect on reading time ($t(33.488) = -2.120$, $p = 0.035$). Using more frequent words lead to significantly shorter reading times ($\mu = 41.99$, $s = 13.00$ seconds) than using less frequent words ($\mu = 53.35$, $s = 19.10$ seconds). For the control group we found no significant effect on reading time ($t(30.283) = -1.828$, $p = 0.077$) (see Table 1).

Word Length. For group D, we found a significant effect on reading time ($t(33.971) = 2.944$, $p = 0.006$). Using shorter words lead to significantly shorter reading times ($\mu = 13.74$, $s = 6.05$ seconds) than using longer words ($\mu = 21.23$, $s = 10.28$ seconds). For the control group, there was no significant effect on reading time ($t(38.569) = 1.957$, p = 0.058$) (Table 2).

Fixation Duration. There was a significant difference between the groups regarding fixation duration ($t(140.111) = 8.209$, $p < 0.001$). Participants with dyslexia had significantly longer fixation times ($\mu = 0.23$, $s = 0.05$ seconds) than the participants without dyslexia ($\mu = 0.17$, $s = 0.03$ seconds).

Word Frequency. For the participants with dyslexia, we found a significant effect on fixation duration ($t(35.741) = -2.150$, $p = 0.038$). Using more frequent words lead to significantly shorter fixation times ($\mu = 0.22$, $s = 0.04$ seconds) than using less frequent words ($\mu = 0.25$, $s = 0.05$ seconds). For the control group, no significant effect on fixation duration was found ($t(37.402) = -1.035$, $p = 0.307$) (see Table 1).

Word Length. We found no significant effect on the fixation duration in group D ($t(40.002) = 0.763$, $p = 0.450$), nor in group N ($t(41.526) = 0.991$, $p = 0.327$) (see Table 2).

Comprehension Score. Participants with dyslexia answered fewer questions correctly ($\mu = 82.74\%$, $s = 28.51\%$) than participants without dyslexia ($\mu = 89.88\%$, $s = 20.21\%$). However, the difference between the groups was not statistically significant ($t(149.593) = -1.873$, $p = 0.063$)).

Word Frequency. We did not found a significant effect on text comprehension for group D ($t(37.764) = 0.370$, $p = 0.714$) nor for group N ($t(33.824) = 1.241$, $p = 0.223$) (see Table 1).

Word Length. For the participants with dyslexia, we found a significant effect on the comprehension score ($t(38.636) = -2.396$, $p = 0.022$). Shorter words led to significantly higher text comprehension ($\mu = 88.64\%$, $s = 26.42\%$) than longer words ($\mu = 65.91\%$, $s = 35.81\%$). For the control group, changing the word length had no significant effect on text comprehension ($t(41.16) = -0.707$, $p = 0.484$) (see Table 2).

6 Discussion

In general, participants without dyslexia read significantly faster and had shorter fixation durations than participants with dyslexia. However, participants with dyslexia read significantly faster and have significantly shorter fixation durations using more frequent words. Using shorter words caused participants with dyslexia to read significantly faster and significantly increased their text comprehension. For the people without dyslexia, no differences in reading time, fixation duration, or text comprehension were found.

Regarding the differences between the groups, our results are consistent with other eye-tracking studies to diagnose dyslexia that found statistical differences between the two populations [14].

Since shorter reading times and fixation durations are associated with better readability, our findings support **H1.1** *A larger number of high frequency words increases readability for people with dyslexia.* The effect of frequency is more pronounced in people with dyslexia than in people without dyslexia, where non significant results were found. Our results are consistent with previous results for both groups. Word frequency was found to have a powerful influence on word recognition tasks for people with [20] or without [33, 44] dyslexia and some reading experiments have demonstrated that readers spend more time looking at low-frequency words than at high-frequency words [20].

In Figure 1 we consider the two visual behaviors that we use as surrogate variables for readability (reading time and fixation duration), visualizing the distribution of the data in such 2D space. A more meaningful feature emerges that can be obtained from a linear combination of both of them.[17] The emerging feature can be interpreted as a readability variable that defines the readability axis. Thus, any readability improvement that moves along the direction determined by the arrow approaches the "ideal situation", which is characterized by target values reached by people without dyslexia in the most favorable conditions (e.g. the behavior on people without dyslexia in texts with higher frequency words). The use of more frequent words bring the average fixation time of group D closer to group N.

Since word frequency had no significant effect on text comprehension, we cannot confirm **H1.2:** *A larger number of high frequency words increases understandability for people with dyslexia*.

Since shorter reading times are associated with better readability, our findings support **H2.1:** *The presence of short words compared to long words increases readability for people with dyslexia*. Although no statistical significance has been found for length regarding fixation duration, we can observe a qualitative improvement in the readability through the use of shorter words since the mean fixation duration for group D also comes closer to the mean fixation duration for group N.

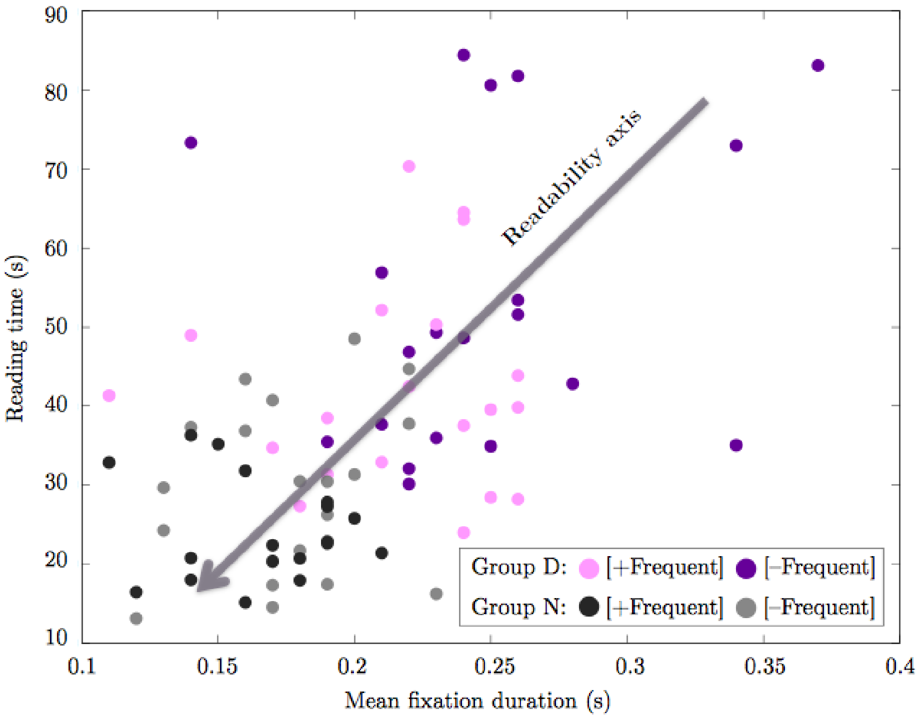


Fig. 1. Readability as a function of reading time and fixation duration

[17] Doing a least squares linear regression we obtain the formula *Reading time* = 173.3 × *Fixation time* + 2.2 seconds (Pearson correlation of 0.51 and $p < 0.001$).

One possible explanation of the lower significance of word length than word frequency is that *Wizard* might have been more difficult to read than *Car*. According to some of the comments of the participants during the open questions, the fact that there was an additive set in front of the noun in the text *Wizard* made it a bit more complicated for further reading. In Spanish, adjectives postponed to nouns are frequent and natural (unmarked syntactic structure for most of the cases), therefore this might be a new variable to take into consideration in further work.

Because participants with dyslexia had a significantly increased text comprehension with texts having shorter words, our findings support **H2.2:** *The presence of short words compared to long words increases understandability for people with dyslexia*.

One of the limitations of our study is that the inferences made from **H2.1** and **H2.2** could be due to the low frequency of long words and not to their length. Through the experimental design we maintained as much as possible both factors —frequency and length— separated for studying both effects independently. However a total dissociation was not possible, as we could not find long words that were more frequent than their shorter synonyms. Although Rayner and Duffy [33] explain the necessity to study both effects separately, such separation between token frequency of linguistic expressions and their length does not exist in natural language [6]. Words [+long] are in average 59.45 less frequent than their correspondent [–long] synonyms. However this frequency ratio is not as high as in the experiment *word frequency* where [+frequent] were 1249.17 times more frequent in average than their [–frequent] synonyms (see Appendix for the frequency ratios for each pair). Since using shorter words usually implies more frequent words, **H1.1** and **H2.1** reinforce each other.

However, the reasons why longer and less frequent words receive longer fixations are different [20], because while less frequent words require more processing, the word-length effect can be ascribed to acuity limitations of the visual system.[18] On the other hand, longer words may imply more fixations instead of longer fixations.

7 Conclusions and Future Work

We tested the effect of word length and frequency on readability and comprehension. Our results show that more frequent words improve readability while shorter words may improve understandability, especially in people with dyslexia. The results also provide evidence about the potential of measuring fixation duration and reading time as a visual behavior surrogate for the readability of a text. The results suggest that people without dyslexia come closer to the ideal reading scenario with shorter fixations and reading time as well as better text comprehension.

These findings can have great impact on interactive systems that rely on text as the main information medium, such as browsers, PDF viewers, or eBook readers. By applying our suggested lexical simplification strategies, namely substituting long and non-frequent words through shorter and frequent synonyms, these systems could

[18] Long words extend beyond the fovea where the acuity is greatest, thus increasing the need for making a fixation and even a re-fixation on a word [20].

make texts easier to read and understand for people with dyslexia. We implemented an automatic method for lexical simplification in Spanish [2] and tried two different interactive strategies with participants with dyslexia [37]. We found that showing simpler synonyms on demand had a better impact than substituting such synonyms. Further improvements on the user interactions are being done and final results will be integrated in *DysWebxia 2.0* [38], a tool that modifies text content and layout for people with dyslexia.

Since there is a correlation between word length and frequency, as frequently used words tend to be shorter, future work needs to investigate potential inter-dependencies between these two factors. We will explore also to what extent lexical simplifications can improve the readability of texts that people with dyslexia read as part of their daily life, such as news, Facebook entries, or Wikipedia articles.

Acknowledgments. We thank Maarten Janssen for facilitating us the dictionary entries and Giorgio Brajnik for his invaluable comments and advise. We thank Mari-Carmen Marcos for her assistance with the eye-tracker hardware and Joaquim Llisterri for his help distributing our experiments announcement among experts. This research was funded by the Spanish Ministry of Education and Science (Grant TIN2009-14560-C03-01) and the FI Catalonian scholarship program.

References

1. Al-Wabil, A., Zaphiris, P., Wilson, S.: Web navigation for individuals with dyslexia: An exploratory study. In: Stephanidis, C. (ed.) HCI 2007. LNCS, vol. 4554, pp. 593–602. Springer, Heidelberg (2007)
2. Bott, S., Rello, L., Drndarevic, B., Saggion, H.: Can Spanish be simpler? LexSiS: Lexical simplification for Spanish. In: Proc. Coling 2012, Mumbay, India (December 2012)
3. Bott, S., Saggion, H.: Text simplification tools for Spanish. In: Proceedings of the Eighth International Conference on Language Resources and Evaluation, LREC 2012, Istanbul, Turkey (May 2012)
4. British Dyslexia Association. Dyslexia style guide (January 2012), `http://www.bdadyslexia.org.uk/`
5. Brunswick, N.: Unimpaired reading development and dyslexia across different languages. In: McDougall, S., de Mornay Davies, P. (eds.) Reading and Dyslexia in Different Orthographies, pp. 131–154. Psychology Press, Hove (2010)
6. Bybee, J.: Phonology and language use, vol. 94. Cambridge Univ. Press (2003)
7. Carrillo, M.S., Alegría, J., Miranda, P., Pérez, S.: Evaluación de la dislexia en la escuela primaria: Prevalencia en español (Evaluation of dyslexia in primary school: The prevalence in Spanish). Escritos de Psicología 4(2), 35–44 (2011)
8. ClaroSoftware (2012), `http://www.clarosoftware.com/index.php?cPath=348`
9. Coltheart, M.: Phonological dyslexia: Past and future issues. Cognitive Neuropsychology 13(6), 749–762 (1996)
10. Coltheart, M., Davelaar, E., Jonasson, T., Besner, D.: Access to the internal lexicon. In: Attention and Performance VI, pp. 535–555 (1977)

11. Cuetos, F., Valle, F.: Modelos de lectura y dislexias (Reading models and dyslexia). Infancia y Aprendizaje (Infancy and Learning) 44, 3–19 (1988)
12. Dixon, M.: Comparative study of disabled vs. Non-disabled evaluators in user-testing: Dyslexia and first year students learning computer programming. In: Stephanidis, C. (ed.) HCI 2007. LNCS, vol. 4554, pp. 647–656. Springer, Heidelberg (2007)
13. Drndarevic, B., Saggion, H.: Towards automatic lexical simplification in Spanish: an empirical study. In: Proc. NAACL HLT 2012 Workshop PITR 2012 (2012)
14. Eden, G., Stein, J., Wood, H., Wood, F.: Differences in eye movements and reading problems in dyslexic and normal children. Vision Research 34(10), 1345–1358 (1994)
15. Ellis, A.: Reading, writing and dyslexia. Erlbaum, London (1984)
16. Evett, L., Brown, D.: Text formats and web design for visually impaired and dyslexic readers-clear text for all. Interacting with Computers 17, 453–472 (2005)
17. Gregor, P., Dickinson, A., Macaffer, A., Andreasen, P.: Seeword: a personal word processing environment for dyslexic computer users. British Journal of Educational Technology 34(3), 341–355 (2003)
18. Gregor, P., Newell, A.F.: An empirical investigation of ways in which some of the problems encountered by some dyslexics may be alleviated using computer techniques. In: Proceedings of the Fourth International ACM Conference on Assistive Technologies, ASSETS 2000, New York, NY, USA, pp. 85–91 (2000)
19. Huenerfauth, M., Feng, L., Elhadad, N.: Comparing evaluation techniques for text readability software for adults with intellectual disabilities. In: Proc. ASSETS 2009, pp. 3–10. ACM (2009)
20. Hyönä, J., Olson, R.: Eye fixation patterns among dyslexic and normal readers: Effects of word length and word frequency. Journal of Experimental Psychology: Learning, Memory, and Cognition 21(6), 1430 (1995)
21. Interagency Commission on Learning Disabilities. Learning Disabilities: A Report to the U.S. Congress. Government Printing Office, Washington DC, U.S. (1987)
22. International Dyslexia Association. Definition of dyslexia (2011), `http://interdys.org/DyslexiaDefinition.htm`; Based in the initial definition of the Research Committee of the Orton Dyslexia Society, former name of the IDA, done in 1994
23. Jiménez, J.E., Guzmán, R., Rodríguez, C., Artiles, C.: Prevalencia de las dificultades específicas de aprendizaje: La dislexia en español (The prevalence of specific learning difficulties: Dyslexia in Spanish). Anales de Psicología (Annals of Psychology) 25(1), 78–85 (2009)
24. Jurafsky, D., Bell, A., Gregory, M., Raymond, W.: Evidence from reduction in lexical production. Frequency and the Emergence of Linguistic Structure 45, 229 (2001)
25. Just, M., Carpenter, P.: A theory of reading: From eye fixations to comprehension. Psychological Review 87, 329–354 (1980)
26. Kanvinde, G., Rello, L., Baeza-Yates, R.: IDEAL: a dyslexic-friendly e-book reader (poster). In: Proc. ASSETS 2012, Boulder, USA. ACM Press (October 2012)
27. Lachmann, T., Van Leeuwen, C.: Different letter-processing strategies in diagnostic subgroups of developmental dyslexia. Cognitive Neuropsychology 25(5), 730–744 (2008)
28. Longman, P.: Longman dictionary of contemporary English. Pearson Longman (2009)
29. McCarthy, J.E., Swierenga, S.J.: What we know about dyslexia and web accessibility: a research review. Universal Access in the Information Society 9, 147–152 (2010)
30. Nation, K., Snowling, M.: Individual differences in contextual facilitation: Evidence from dyslexia and poor reading comprehension. Child Development 69(4), 996–1011 (1998)

31. Patterson, K., Marshall, J., Coltheart, M.: Surface dyslexia: Neuropsychological and cognitive studies of phonological reading. Lawrence Erlbaum Associates, London (1985)
32. Pérez, M., Alameda Bailén, J., Cuetos Vega, F.: Frecuencia, longitud y vecindad ortográfica de las palabras de 3 a 16 letras del diccionario de la lengua española (Frequency, length and orthographical neighborhood of the words from 3 to 16 letters of the dictionary of the Spanish language – RAE 1992). REMA: Revista Electrónica de Metodología Aplicada (Electronic Journal of Applied Methodology) 8(2), 1–10 (2003)
33. Rayner, K., Duffy, S.: Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. Memory & Cognition 14(3), 191–201 (1986)
34. Real Academia Española. Diccionario de la lengua española (Dictionary of the Spanish language). Espasa-Calpe, Madrid, 22nd edn (2001)
35. Rello, L., Baeza-Yates, R.: Lexical quality as a proxy for web text understandability (poster). In: Proc. WWW 2012, Lyon, France (2012)
36. Rello, L., Baeza-Yates, R.: The presence of English and Spanish dyslexia in the Web. New Review of Hypermedia and Multimedia 8, 131–158 (2012)
37. Rello, L., Baeza-Yates, R., Bott, S., Saggion, H.: Simplify or help? Text simplification strategies for people with dyslexia. In: Proc. W4A 2013, Rio de Janeiro, Brazil (2013)
38. Rello, L., Baeza-Yates, R., Saggion, H., Bott, S., Carlini, R., Bayarri, C., Gòrriz, A., Gupta, S., Kanvinde, G., Topac, V.: Dyswebxia 2.0!: Accessible text for people with dyslexia (demo). In: Proc. W4A 2013, The Paciello Group Web Accessibility Challenge, Rio de Janeiro, Brazil (2013)
39. Rello, L., Baeza-Yates, R., Saggion, H., Graells, E.: Graphical schemes may improve readability but not understandability for people with dyslexia. In: Proceedings of the NAACL HLT 2012 Workshop Predicting and Improving Text Readability for Target Reader Populations, PITR 2012 (2012)
40. Rello, L., Bautista, S., Baeza-Yates, R., Gervás, P., Hervás, R., Saggion, H.: One half or 50%? An eye-tracking study of number representation readability. In: Kotzé, P., Marsden, G., Lindgaard, G., Wesson, J., Winckler, M. (eds.) INTERACT 2013, Part IV. LNCS, vol. 8120, pp. 229–245. Springer, Heidelberg (2013)
41. Rello, L., Bayarri, C., Gorriz, A.: What is wrong with this word? Dyseggxia: a game for children with dyslexia (demo). In: Proc. ASSETS 2012, Boulder, USA. ACM Press (October 2012)
42. Rello, L., Kanvinde, G., Baeza-Yates, R.: Layout guidelines for web text and a web service to improve accessibility for dyslexics. In: Proc. W4A 2012, Lyon, France. ACM Press (2012)
43. Rüsseler, J., Probst, S., Johannes, S., Münte, T.: Recognition memory for high-and low-frequency words in adult normal and dyslexic readers: an event-related brain potential study. Journal of Clinical and Experimental Neuropsychology 25(6), 815–829 (2003)
44. Sereno, S., Rayner, K.: Measuring word recognition in reading: eye movements and event-related potentials. Trends in Cognitive Sciences 7(11), 489–493 (2003)
45. Seymour, P., Aro, M., Erskine, J.: Foundation literacy acquisition in European orthographies. British Journal of Psychology 94(2), 143–174 (2003)
46. Simmons, F., Singleton, C.: The reading comprehension abilities of dyslexic students in higher education. Dyslexia 6(3), 178–192 (2000)
47. Stanley, G., Smith, G., Howell, E.: Eye-movements and sequential tracking in dyslexic and control children. British Journal of Psychology 74(2), 181–187 (1983)
48. Tobii Technology. Product description Tobii 50 Series (2005)
49. Topac, V.: The development of a text customization tool for existing web sites. In: Text Customization for Readability Symposium (November 2012)

50. Vellutino, F., Fletcher, J., Snowling, M., Scanlon, D.: Specific reading disability (dyslexia): What have we learned in the past four decades? Journal of Child Psychology and Psychiatry 45(1), 2–40 (2004)
51. Williams, S., Reiter, E., Osman, L.: Experiments with discourse-level choices and readability. In: Proc. ENLG 2003, Budapest, Hungary (2003)
52. Wimmer, H.: Characteristics of developmental dyslexia in a regular writing system. Applied Psycholinguistics 14(1) (1993)

Appendix: Synonyms Pairs

The list of the unique pairs of synonyms used in the experiment *word-frequency* is below. The frequency ratio appears in parenthesis after the [±frequent].

Pakistan [+frequent]	[−frequent]	*Somalia* [+frequent]	[−frequent]
ataques	refriegas (474)	personas	individuos (26)
sequía	agostamiento (903)	casas	moradas (173)
entrega	avituallamiento (787)	inundaciones	aluviones (123)
casa	morada (67)	lluvias	diluvios (195)
personas	individuos (26.3)	verano	canícula (21,140)
ciudades	urbes (106)	desastre	hecatombe (31)
ejército	hueste (209)	parte	porción (58)
rebeldes	insubordinados (511)	país	territorio (6)
ciudad	capitalidad (484)	estado	tesitura (741)
producción	obtención (39)	comunidad	colectividad (89)
alimentos	sustentos (968)	generosidad	dadivosidad (285)
precios	valías (5,897)	velocidad	apresuramiento (491)
gente	muchedumbre (107)	necesidades	indigencias (19)
productos	manufacturas (133)		
mercado	baratillo (888)		

The list of the unique pairs of synonyms used in the experiment *word-length* is shown below. The frequency ratio appears in parenthesis after the [+long] word. The length ratio is shown after the frequency ratio.

Car [−long]	[+long]	*Wizard* [−long]	[+long]
huída	escabullida (81; 2.2)	piso	apartamento (2; 2.75)
parking	estacionamiento (15; 2.14)	mago	prestidigitador (87; 3.75)
pavor	sobrecogimiento (13; 2.8)	raro	estrambótico (10; 3.5)
auto	autómovil (65; 2.25)	aumento	acrecentamiento (6; 2.14)
raro	extravagante (7; 3)		
cara	semblante (276; 2.25)		
pálida	emblanquecido (93; 2.16)		

TicQR: Flexible, Lightweight Linking of Paper and Digital Content Using Mobile Phones

Jennifer Pearson[1], Simon Robinson[1], George Buchanan[2], and Matt Jones[1]

[1] FIT Lab, Swansea University
[2] Centre for HCI Design, City University
{j.pearson,s.n.w.robinson,matt.jones}@swan.ac.uk

Abstract. In this paper we introduce TicQR – a photo-based checkbox-enabled interface which bridges the physical and digital document domains, allowing automatic download or processing of useful data from paper documents. There is a long demonstrated need for people to be able to connect between printed material and digital information and services. By using a combination of image recognition and QR codes we are able to detect user marks on paper documents via a single photograph taken with a standard smart phone. This information can then be used to access the equivalent digital content, save contacts or URLs, or even order goods directly from local retailers.

Keywords: Paper documents, tick boxes, QR Codes.

1 Introduction

Digital services now permeate every aspect of our lives. However, printed materials predominate in many everyday situations, and are commonly used to gather or communicate vital information. But paper itself does not connect directly to digital services; for that reason many have sought to create links to the digital domain via a range of technologies. Most existing solutions require specialist equipment to connect paper and digital content, or suffer from relatively coarse information granularity – for example, connecting a whole page or document to a single digital item. While these approaches do indeed connect paper to digital, we argue that it would be preferable to be able to use more commonplace mobile technologies, and at the same time be able to provide finer-grained information, even considering specific marks or annotations by the user on the paper document.

To achieve this goal, we present a novel technologically and interactively lightweight approach that lets readers extract information from a printed document. Our approach – TicQR, allows people to tick boxes to select interesting sections directly on paper, and photograph the page using a standard cameraphone. Given both the document's identity and the specific selections made, TicQR can then undertake a variety of actions depending on how the document's designer has configured the support for their text. We thus provide a rich opportunity for interaction that embraces both print and digital media, but requires no additional hardware or specialist paper.

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 220–228, 2013.

2 Background

Paper documents have been around for millennia, and despite the ever-growing popularity of digital reading, paper remains ubiquitous. The physicality of paper offers affordances that computer screens or eReaders do not – such as folding and scribbling, for example. Globally, for reasons such as cost, physicality, aesthetics and utility, paper remains a desirable medium in many cases [8]. Yet the digital medium has many advantages, including support for searches across content, links between items, and its physical space benefits. Mixed approaches offer a combination of the strengths of each form.

The TicQR design aims to bridge the physical-digital document gap in a manner that is minimally intrusive to the user. The system allows people to read a paper document, mark which parts they are interested in, and archive them digitally by taking a single photograph using a standard smartphone handset (see Fig. 1). This approach allows multiple 'clippings' on a page to be recognised and interpreted simultaneously, then presented to the user for browsing or later use. We envisage many scenarios where TicQR could be useful, some of which are illustrated later in Fig. 2.

Fig. 1. Using TicQR to interact with a newspaper. Left: reading a newspaper, finding and then ticking interesting content. Right: browsing saved items at a later time.

2.1 Related Work

Research into connecting physical and digital media is well established. An early example is the DigitalDesk [10], which combined a desk, and a camera and projectors mounted above, to provide digital augmentation of paper documents. A paper document was placed in a fixed location on the desk, and text was extracted using OCR, allowing cut-and-paste from paper into a digital medium. The DigitalDesk showed the potential benefits of linking print and electronic formats, but required a precisely controlled, calibrated environment.

Digital paper and pens are a more direct combination of printed and electronic media, typified by Anoto's digital pen technology[1]. PaperPoint used Anoto to support real-time annotation of on-screen presentations using a printout of the slideshow [9]. PaperLink [1] used a custom video pen to associate annotations with digital content. In contrast, Papiercraft [4] used Anoto ink markup of the printout of a digital document to manipulate the digital original. Using Anoto and similar techniques can allow users to trigger specific digital actions from ink on paper documents. However, augmented, patterned paper is required, and a dedicated battery-powered pen must be used to annotate. Printing must also be carefully controlled so that the right layout is associated with each piece of paper.

The *a-book* [5] also uses dedicated equipment to link paper and digital items. The system makes use of a graphics tablet placed under a paper document to allow manipulation (e.g., linking, searching) of handwritten content via a second display. Each of these systems demonstrates the potential for rich interaction and effective usability when using dedicated hardware. However, the associated costs and constraints impact on both developers and users. In response to such approaches, researchers have proposed solutions that use simpler, commonplace technologies, therefore allowing more individuals, organisations and communities to use such techniques.

One common contemporary example is QR codes, which allow direct links to digital material via a standard mobile phone camera. Scanning a QR code typically provides a single piece of information, however – for example, a single contact or a web link. If multiple pieces of information are required, then the user will need to scan multiple codes, or follow web links. Previous solutions to this issue have used one code or other identifier per page of media and then automatically downloaded all information from the page[2]. This approach could easily result in downloading unwanted information, and cause higher user task loads as this content is filtered out.

It is well documented that paper is the preferred medium for annotation (e.g., [6,7]) and other small-scale information work. Our approach takes advantage of this fact, and allows users to mark on the document itself to select items. The mScan project [2] also uses camera phones and users' marks on paper, but to scan a specific, preset form with multiple-choice bubble fields, that is placed in a known position and orientation. While encouragingly reliable, the mScan format only supports very specific marks for data gathering. A dedicated stand is also required to fix the location of a form for analysis, much as with the DigitalDesk (cf. [10]) or a standard flatbed scanner.

To summarise, in contrast to dedicated equipment approaches, our technique uses QR codes to identify documents, and requires only a standard mobile phone camera for image processing. Our approach can also be used on-the-go, allowing the user to browse their clippings in a more 'laid-back' [3] manner than previous QR-powered document scanning designs. Finally, like PapierCraft, Paperpoint and the DigitalDesk, the TicQR technique allows richer physical-digital interaction with paper documents.

---

[1] See: `anoto.com`.

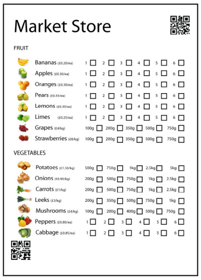
[2] For example, see `kooaba-shortcut.com` or augmentation examples such as `aurasma.com`.

3 The TicQR System

The TicQR system offers richer experiences with physical documents by allowing people to tick items they are interested in on the paper itself. Our approach uses a combination of precisely-placed QR codes and image recognition to enable the system to determine both which document has been scanned; and, which checkboxes within the document have been selected. Fig. 2 illustrates several example usage scenarios and corresponding QR code-augmented paper documents that we have created

Scenario 1: Newspapers

Sally is reading a newspaper while waiting for a bus. She has read several interesting news stories and has also spotted an advert for a book she would like to buy as a gift. Using a regular pen, Sally ticks the useful news stories and book advert on the newspaper and takes a picture of the entire page using her mobile phone camera. Putting her phone away and leaving the newspaper on the seat, she heads on her way. Later that evening, she opens the TicQR app and checks her clippings library. Inside, she finds that a digital copy of the news stories and advert she ticked have been automatically downloaded and added to her library for easy browsing and manipulation. Sally now has an archive of news clippings and a link to the online retailer that was advertising the book.



Scenario 2: Local Shopping

Alex is planning a dinner party and needs to purchase several fresh ingredients. Earlier that week, he remembers an advert that was posted through his letter box from his local store. This flyer consisted of a list of the products available with tick boxes alongside them. Alex reads through the paper list and ticks all the items he requires. He gets out his phone, takes a picture of the flyer and goes to get ready. A short while later, Alex walks to his local shop where his basket of goods is there waiting to be paid for and collected.

Scenario 3: Take-out

It's Saturday night and Nathan and his friends want to order take-out for dinner. Deciding on Chinese, the group sift through the pile of take out menus they frequently get posted through their door and select one augmented with TicQR. They proceed to tick the items and quantities they want from the menu and complete the order by taking a photograph with the TicQR app. Using the address information Nathan added when he downloaded the application, the local take-away delivers the food to his door.

Scenario 4: Tailor Made Grocery Lists

Jill does one large online supermarket delivery every week. The interface of the supermarket website allows her to save her recurring items in a list for easy reordering, and has recently added TicQR functionality. The website allows her to print a TicQR enabled list of past items which she then sticks on her fridge. When something she regularly uses runs out, she immediately ticks the box on the paper list so she does not forget later. At the end of the week, Jill takes a picture of the list which automatically adds the items she requires to her online order.

Fig. 2. Potential usage scenarios for the TicQR system. For scenarios 1–3, a TicQR-augmented paper document is also shown. Scenarios 1 and 2 were evaluated in our study.

for the TicQR system. To retrieve digital content from a paper document, the user first positions their cameraphone to take a photo of the item.

The TicQR application automatically takes a photograph when it detects two QR codes which are positioned at opposite corners of the object – one at the bottom left and another at the top right (see Fig. 2). The bottom left code contains the identifier for the item (e.g., the issue and page number of a newspaper), and the top right code is used for image alignment and coordinate calibration. Using the identifier from the QR code, the application queries an online database to retrieve the item's metadata, which contains the approximate coordinate of each of its checkboxes.

The current setup requires that the server side (for example, a newspaper publisher) provides a web service that the TicQr application can query to determine checkbox positions, as well as the content that should be attributed to the selection of each checkbox. If at the time the photo was taken there is no internet connectivity, the application simply saves the photo and performs these operations when a connection is available.

When the item's metadata has been downloaded, the image is then processed to determine if any of its boxes have been marked. When a list of marked boxes has been generated, the application performs actions based on the specific items selected (see Fig. 2 for examples).[3] Any downloaded content is saved locally in the application's clippings repository for later use (see Fig. 1 and Fig. 3). All processing happens in the background of the application – we currently use the OpenCV[4] library for checkbox processing, which takes less than one second to complete.

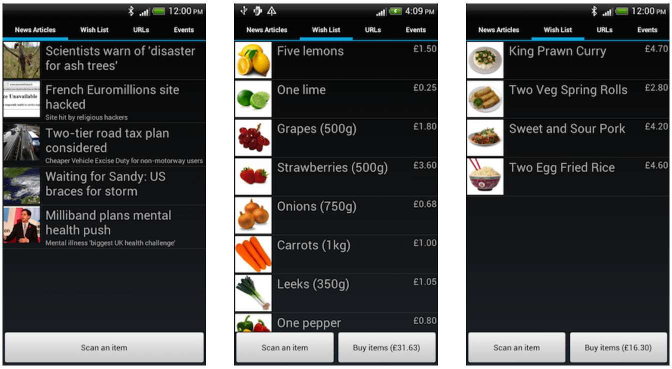


Fig. 3. Example TicQR interfaces. Left: a selection of newspaper clippings; Middle: a grocery order from a local market; Right: a sample order from a take-away menu.

4 Evaluation

We performed a lab study in order to help understand both the TicQR system's recognition accuracy and its reception by users. 33 participants (16M 17F, aged 18–53) took part in individual 30 min trials. The metrics used were the accuracy rate of tick box detection, and qualitative data in the form of interview responses.

[3] A video illustrating the technique is also available – see: `goo.gl/QBucU`

[4] See: `opencv.org`

4.1 Prototype Media

We selected two types of prototype media to use in the evaluation – a newspaper (as described in Scenario 1 in Fig. 2), and a shop order form (as described in Scenario 2 in Fig. 2). The newspaper was chosen as its form offers a wide selection of different clipping types (e.g., articles, products, events), and typically includes several types on a single page. The mock-up newspaper used in the study included five news stories, one event advert, one product advert, one URL and one contact. The shop order form used was designed as an extensive product list, which allows consumers the freedom to order specific quantities of a particular product, but also has the benefit of "stress-testing" the recognition system with a closely-spaced grid of checkboxes.

4.2 Procedure

At the start of each session the participant was welcomed and led through an ethically-approved consent procedure. Following this, the system was demonstrated once to the participant, and they scanned an example pre-ticked document as a form of training. We explained the two use case scenarios that would be examined in the study to provide participants with a context in which the system could be used. Participants were then given three identical blank copies of each type of prototype document (i.e., six documents in total), and asked to mark a selection of boxes of their choice. After ticking any number of boxes on a form, participants used the TicQR application to take a photograph of the document and either navigate the clippings library (newspaper), or order the list of marked products (order form). Participants were not told which boxes to mark, nor were they instructed on *how* to mark them (e.g., whether to tick, cross, scribble etc.).

We recorded the accuracy rate of the system's checkbox recognition based on the total number of boxes marked, and the number of false positives (i.e., un-ticked boxes detected as ticked) and false negatives (i.e., ticked boxes not detected). To conclude the study we conducted a short semi-structured interview to gather participants' existing document markup behaviours and their opinions of the TicQR system. The participant was then given a gift voucher as token of our appreciation.

4.3 Results and Discussion

Robustness. A total of 1324 ticked boxes were recorded over all participants in the study (an average of seven per document). The overall checkbox recognition accuracy was 98.2%. Of the incorrectly recognised checkboxes, 17 were false positives (1.28%) and 7 were false negatives (0.52%). These mis-recognised boxes were rarely due to problems with the algorithm used to detect user marks, however. The majority of false positives were caused by factors which were out of our control, including the way in which participants marked up the paper documents – for example, when the end of a tick from one box inadvertently passed through another (see Fig. 4 (a) and (b) for examples taken from actual study participants).

Another minor issue that caused some of the false positive results during the study involved the way in which participants took the initial photograph. That is, if the photograph was taken when the image was slightly rotated, this could mean that some checkboxes at the edges of the document were only partially included in the photograph. The checkbox recognition algorithm used for the study ignored boxes which were completely outside the image, but detected boxes partially outside the image as ticked (we have since refined the algorithm to mark partial boxes as unticked). Of the 17 false positives observed, 9 were as a result of participants' marks running through multiple boxes, 4 were caused by the corners of the box being occluded due to participants rotating the camera, and a further 4 were true false positives.

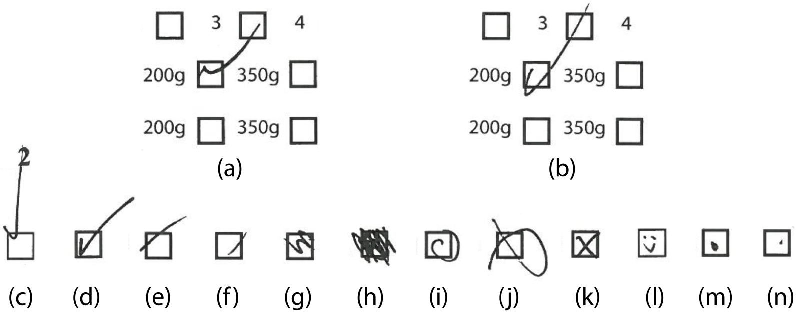


Fig. 4. Sample tick marks made by study participants. Top: participants occasionally ticked multiple boxes with one tick. All other marks except (n) were recognised correctly.

Several of the false negatives recorded were due to participants intentionally trying to fool the system. For example, the mark shown in Fig. 4 (m) was drawn by a participant who stated immediately after drawing it "let's see how well it gets this one", while another with a similar marking stated "I want to see if I can break the app". That said, however, the system was able to pick up on the majority of users' marks, regardless of how they were drawn. Fig. 4 (c–n) illustrates a representative selection of marks observed during the study, ranging from ticks, to lines, to crosses, scribbles and even a smiley face. With the exception of the small dot shown in Fig. 4 (n), all other marks shown were recognised by the TicQR system.

The majority of inaccuracies within the study were recorded on the order form rather than the newspaper, and in fact all false positives caused by rotation issues or user marks running through multiple boxes were recorded on the order form. We attribute this to the close proximity of the checkboxes to one another and to the edge of the form. This document was designed to test the recognition of the system, and is therefore not necessarily an accurate representation of a real order form. However, to reduce these issues, checkboxes could be placed further apart from each other, and away from the top left and bottom right corners of the document.

Subjective Responses. All participants spoke favourably about the TicQR technique, giving an average score of 8.1 out of 10 (lowest 6, highest 10) for the usefulness of the system. Comments made by participants on the topic included: "I do love my shortcuts so anything to make my life easier is great" "it's very simple and intuitive – I really 'got' it"; "it [TicQR] is great – I often take photos of documents myself to look back at later, so an interactive photo to source things you're interested in is very efficient"; and, "The app does all the legwork for you – you don't have to trawl the internet to find things".

A majority of participants (88%; 29 of 33) reported finding themselves in a situation where they had read something on a physical document and later wanted to locate a digital version. The most common reason for this was to allow them to email, search or archive the digital copy. More interestingly, however, when asked if they did actually manage to locate a digital copy only 21% (6 participants) of these 29 had done so without some level of difficulty. The remainder had either not managed to locate a digital copy at all (14%; 4 participants), only managed to locate it on some occasions (44%; 13 participants), or had found the digital copy only after some time or effort (21%; 6 participants). Some even described alternatives for locating a proper digital version such as taking a photograph of the physical document for later reference (21%; 6 participants).

5 Conclusions and Future Work

The TicQR technique enables automatic downloads of user-selected digital content from a paper document. While previous mobile phone-based methods required either the use of a search engine to locate related digital material, or the download of a complete package of digital content via a single QR code, TicQR affords the freedom to select only the information required via marks on the physical medium itself, and downloads selected content automatically in a single step. In contrast to previous interaction techniques, TicQR uses a standard cameraphone, which can be exploited with a much lower investment by individuals and small organisations.

Our prototype and user study demonstrate that the TicQR technique is practical and effective both technologically and in terms of usability. We have evaluated two concrete use cases for the TicQR technique, and provided several supplementary scenarios where it would also be beneficial. TicQR is thus a powerful and affordable tool for delivering richer interactive services than many existing techniques.

Our next step is to collaborate with a local produce market to provide a TicQR augmented ordering system. We hope that a longitudinal evaluation such as this will provide a fuller picture of the benefits of the system, and show how the combination of physical documents and digital ordering can provide richer user interaction.

Acknowledgements. This work was funded by EPSRC grant number EP/J000604/2.

References

1. Arai, T., Aust, D., Hudson, S.: PaperLink: a technique for hyperlinking from real paper to electronic content. In: Proc. CHI 1997, pp. 327–334. ACM (1997)
2. Dell, N., Breit, N., Chaluco, T., Crawford, J., Borriello, G.: Digitizing paper forms with mobile imaging technologies. In: Proc. DEV 2012. ACM (2012)
3. Jones, M., Jain, P., Buchanan, G., Marsden, G.: Using a mobile device to vary the pace of search. In: Chittaro, L. (ed.) Mobile HCI 2003. LNCS, vol. 2795, pp. 390–394. Springer, Heidelberg (2003)
4. Liao, C., Guimbretiére, F., Hinckley, K.: PapierCraft: a command system for interactive paper. In: Proc. UIST 2005, pp. 241–244. ACM (2005)

5. Mackay, W., Pothier, G., Letondal, C., Bøegh, K., Sørensen, H.: The missing link: augmenting biology laboratory notebooks. In: Proc. UIST 2002, pp. 41–50. ACM (2002)
6. Marshall, C.: Reading and Writing the Electronic Book. Morgan and Claypool (2009)
7. O'Hara, K., Sellen, A.: A comparison of reading paper and on-line documents. In: Proc. CHI 1997, pp. 335–342. ACM (1997)
8. Sellen, A., Harper, R.: The Myth of the Paperless Office. MIT Press (2003)
9. Signer, B., Norrie, M.: PaperPoint: a paper-based presentation and interactive paper prototyping tool. In: Proc. TEI 2007, pp. 57–64. ACM (2007)
10. Wellner, P.: Interacting with paper on the DigitalDesk. Commun. ACM 36(7), 87–96 (1993)

One Half or 50%? An Eye-Tracking Study of Number Representation Readability

Luz Rello[1,2], Susana Bautista[3], Ricardo Baeza-Yates[1,4], Pablo Gervás[3], Raquel Hervás[3], and Horacio Saggion[2]

[1] Web Research Group, Universitat Pompeu Fabra, Barcelona
[2] Natural Language Processing Group, Universitat Pompeu Fabra, Barcelona
[3] Natural Interaction based on Language, Universidad Complutense de Madrid
[4] Yahoo! Labs, Barcelona
{luzrello,rbaeza}@acm.org, {subautis,raquelhb}@fdi.ucm.es, pgervas@sip.ucm.es, horacio.saggion@upf.edu

Abstract. Are numbers expressed as digits easier to read and understand than written with letters? What about fractions and percentages? Exact or rounded values? We present an eye-tracking study that attempts to answer these questions for Spanish, using fixation and reading time to measure readability as well as comprehension questions to score understandability. We find that digits are faster to read but do not help comprehension. Fractions help understandability while percentages help readability. No significant results were found concerning the influence of rounding. Our experiments were performed by 72 persons, half of them with dyslexia. To the best of our knowledge, this is the first study that addresses the cognitive load of number representation in any language, even more for people with dyslexia.

Keywords: textual accessibility, dyslexia, user testing, eye-tracking, readability, comprehension, number representation.

1 Introduction

Worldwide, around 15-20% of the population has a language-based disability; where 70-80% of it is likely dyslexic [27]. Some of the these disabilities make more difficult the comprehension of texts written for a generic public. For this reason, United Nations [36] recommend that all public information services and documents should be accessible to the widest possible readership. Given this, there are different initiatives that propose guidelines to help rewriting a text to make it more comprehensive. Some of them are *Plain Language*[1] and the *European Guidelines for the Production of Easy-to-Read Information* [21].

In addition, a large percentage of information expressed in daily news or reports contain numerical expressions (economical statistics, demographic data, etc.), but

[1] http://www.plainlanguage.gov/

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 229–245, 2013.

many people have problems understanding complex expressions, including non-native speakers, and people with limited education or some kind of social or cognitive disability, such as dyslexia. Numerical information can have different representations such as: using digits or words, rounded numbers or decimals, fractions instead of percentages, etc. According to cognitive studies, numbers in a text are processed in a different way than words [13], and the presence of numbers in the text impacts the reading process [34].

Dyslexia is a neurological reading disability characterized by difficulties with accurate and/or fluent word recognition and by poor spelling and decoding abilities [50]. Depending on the language, the estimation of the prevalence of dyslexia varies from 10-17.5% for the population in the U.S.A. [26] to 8.6-11% for the Spanish speaking population [8, 29, 42]. People with dyslexia find problems to recognize and recollect not only letters but also numbers [11, 37]. Although dyscalculia[2] and dyslexia are two different disabilities, they are comorbid [33][3] and people with dyslexia are more likely to have Math learning difficulties [32]. For this reason we are particularly interested in studying how the complexity of numerical expressions affects the comprehension of a text.

In this context, the main goal of this paper is to study how number representation impacts text readability and understandability for native Spanish speakers with or without dyslexia. Readability refers to the legibility of a text, that is, the ease with which text can be read, while understandability refers to comprehensibility, the ease with which text can be understood. With this goal in mind, we conducted three experiments with 72 persons (36 with dyslexia) using eye-tracking and comprehension questionnaires. From our results we can quantify the impact of numerical expressions in the reading process for people with or without dyslexia, and it is possible to apply this information to the adaptation of numerical information so texts are more accessible to the widest number of readers.

To the best of our knowledge, this is the first time that numerical representations are measured in terms of readability and comprehension using our methodology. Therefore, this paper presents three main contributions:

- The first analysis of how numerical information impacts text readability and understandability for people with or without dyslexia using a methodology that includes eye-tracking and comprehension questionnaires.
- Numerical information represented as digits improve readability for people with dyslexia.
- Numerical information represented as percentages improve readability for people with dyslexia.

The rest of the paper is organized as follows. Section 2 covers the related work. In Section 3 we present the details of the experimental methodology. Section 4 presents

[2] A specific learning disability involving innate difficulty in learning or comprehending arithmetic. It is akin to dyslexia and includes difficulty in understanding numbers, learning how to manipulate numbers, learning mathematical facts, and a number of other related symptoms [7].

[3] Comorbidity indicates a medical condition (in this case dyscalculia) existing simultaneously but independently with another condition (dyslexia).

our results and Section 5 discusses them. Conclusions and future work are drawn in Section 6. Finally, in the Appendix, the data used are presented.

2 Related Work

Experimental psychology and cognitive neuropsychology have dealt with the study of number processing and calculation over the last two decades. Many researchers have studied the cognitive processes that are responsible for number processing and calculation, with the goal of contributing to the improvement of teaching and learning processes. For example, [24, 46] present findings about how the frequency of use of a word or number is an influential variable in the reading process. In addition, it seems that numerical expressions most frequently used require less time for recognition.

Researchers in psychophysics [39] have also studied various aspects of reading: impact of text context on readability and eye movements during reading, among others. When numerical expressions are expressed in digits there is a faster access to its semantic representation than when expressed in words [12, 18, 19, 28].

From experimental psychology there is evidence regarding the importance of the frequency of use of numerical expressions. For example, Brysbaert [6] investigated number processing by looking at reading time using eye-tracking and showing that subjects have longer eye fixations with more frequently used numbers.

Overall, it is a generally accepted hypothesis that the probability of making a fixation on a particular linguistic unit is determined by the perceptual or informational relevance of the unit, and the degree of difficulty of processing required for its identification. Due to the quantitative nature of this measure, it is not easy to unravel what are the specific reasons justifying the preference for fixations on certain kinds of linguistic units. Thus, the visual processing of words has been shown to be affected significantly by factors such as length, frequency, type of vocabulary, predictability, or word ambiguity [40].

There is an extensive body of knowledge that takes into consideration eye movements of readers with dyslexia using eye-tracking. While some found differences among people with or without dyslexia [16], others did not [48]. Hyönä *et al.* [25] studies the effect of word length and word frequency in relation with eye fixation patterns. Their results show that low frequency and long words present longer gaze durations and more re-inspections although in [43] only more frequent words presented significantly shorter fixation durations. However, we found no previous work that investigates numerical representations in readers with dyslexia using eye-tracking, nor for people in general. In fact, it is the first user evaluation of numerical representations.

Currently there is plenty of automatic text simplification research based on cognitive aspects. The main objectives are to identify simplification operations that can be applied to adapt a text using some kind of automatic means. Most of the text simplification approaches disregard the treatment of numerical expressions [5, 9] except from [4] that is a numerical expressions simplification system designed on the basis of corpus analyses [2, 15]. However, to the extent of our knowledge there are no user evaluations regarding the impact of simplifying numerical expressions as the one we present in this paper.

3 Methodology

We designed three different experiments to study the effect of different representations of numerical expressions with respect to readability and understandability. In the experiments, 72 participants (36 with dyslexia) had to read several texts in Spanish with different representations of numerical expressions. We tested the following hypotheses:

- H1.1: Readability will increase if digits are used instead of words for representing numerical expressions.
- H1.2: Understandability will increase if digits are used instead of words for representing numerical expressions.
- H2.1: Readability will increase if rounded numerical expressions are used instead of unrounded expressions (with decimals).
- H2.2: Understandability will increase if rounded numerical expressions are used instead of unrounded expressions (with decimals).
- H3.1: Readability will increase if numerical expressions are expressed in percentages instead of fractions.
- H3.2: Understandability will increase if numerical expressions are expressed in percentages instead of fractions.

3.1 Design

In the first experiment, *Digits vs. Words*, the independent variable [±digits] had two levels: [+digit] denotes the condition where numbers in the text were written in digits, *i.e. 22*, and [–digit] denotes the condition where numbers were written using words, *i.e. veintidos ('twenty two').*

In the second experiment, *Rounding vs. Decimals*, the independent variable [±round] had again two levels: [+round] denotes the condition where numbers were rounded, without decimals *i.e. 19* and [–round] denotes the condition where numbers in the text were written with decimals *i.e. 19.45*.

In the third experiment, *Percentages vs. Fractions*, the independent variable [±percentage] also had two levels: [+percentage] denotes the condition where numbers in the text were written using percentages, *i.e. 25%* and [–percentage] denotes the condition where numbers in the text were written using fractions *i.e. 1/4.*

The experiment followed a within-subjects design, so every participant contributed to each of the conditions in the experiments. The order of conditions was counterbalanced to cancel out sequence effects. To measure understandability and readability, we measured fixation duration and the correct answers from the questionnaires, respectively. Sometimes both terms, *readability* and *understandability*, have been used interchangeably. However, previous research with people with dyslexia has shown that both concepts need to be taken into consideration separately. For instance, in [41] comprehension has been found to be independent of readability for people with dyslexia, while text readability can be used as an indicator of comprehension for people without dyslexia. Therefore, in this study we distinguish between readability and understandability.

Fixation Duration. When reading a text, the eye does not move contiguously over the text, but alternates saccades and visual fixations, i.e. jumps in short steps and rests on parts of the text. *Fixation duration* denotes how long the eye rests still on a single place of the text. Fixation duration has been shown to be a valid indicator of readability. According to [30, 40, 47], shorter fixations are associated with better readability while longer fixations can indicate that processing loads are greater. Hence, we use fixation duration across the whole text as measure to quantify readability.

Correct Answers. To measure text comprehension we used questionnaires, one per text. We used multiple-choice questions with three possible choices, one correct and two incorrect. From these answers, we computed the percentage of correct answers, where the correct choice scored 100% and the others 0%.

3.2 Participants

Seventy-two Spanish native speakers undertook the experiments, 36 without dyslexia (group N) and 36 with dyslexia (group D). Their ages ranged from 16 to 50, with a mean age of 26.94 for group N and 23.38 for group D. All the participants were asked to bring their dyslexia diagnosis to the experiment. Except from 3 participants with dyslexia and 2 without who had no higher education, the rest were attending school or high school (17 participants with dyslexia and 13 participants without dyslexia) or were studying or had already finished university degrees (16 participants with dyslexia and 21 participants without dyslexia).

3.3 Materials

To study the effects of numerical expressions we need to study target numerical expressions in context, *i.e.* as part of a text. To isolate the effects of these variables, the texts need to be comparable in complexity. In this section, we describe how we designed the texts and the target numerical expressions that were used in this study.

Base Texts. As basis for our manipulations, we created six texts with an average length of 62.33 words. To meet the comparability requirements among the texts belonging to the same experiment, we took into account the parameters that different complexity measures take into consideration [14]. Next, we present the characteristics shared by the texts in each experiment:

(a) They have the same number of target words: Seven [±digits] numerical expressions pairs for each of the texts in *Digits vs. Words*, seven [±round] expressions pairs for each of the texts in *Rounding vs. Decimals*, and four [±percentages] target numerical expressions in *Percentages vs. Fractions*.

(b) They share the same genre and are about similar topics: fast food ingredients.
(c) They contain the same number of sentences: three sentences in *Percentages vs. Fractions* and four sentences in the other two experiments.
(d) The base texts of each experiment have the same number of words per text (39 words in *Percentages vs. Fractions* and 63 words for the other two experiments).
(e) All the texts have a similar word length average ranging from 4.88 to 5.24 letters.
(f) They do not contain named entities, foreign words, or numerical acronyms.

The presentation of the text has an effect on reading speed so we used the same layout for all the texts. We chose a recommended font type for reading in a screen, sans serif arial, a large size (20 points) and 62 characters/column and unjustified text and recommended color and brightness contrast using a black font with crème background [1, 45].[4]

Target Numerical Expressions. For the numerical expressions we used the following criteria:

(a) We chose pairs of numerical expressions denoting the same number because the frequency of the number expressed has an impact on its processing time [24, 46]. Therefore the same numerical expressions with different representation were chosen for the three experiments.
(b) We did not include ambiguous numerical expressions because they require more processing than unambiguous ones [40]. For instance, depending on the context *cien ('hundred')* could also mean "very fast" in Spanish i.*e. Estoy trabajando a cien ('I am working very fast')* and *Las manzanas están a cien ('The apples cost one hundred').*
(c) We did not used orthographically similar numerical expressions, (i.e. mirror numbers '6' and '9') in the same text since such representations can be a major difficulty for people with dyslexia [17].
(d) In the experiment *Digits vs. Words*, we selected different numerical expressions, *i.e.,* with decimals, rounded, two or three digits and percentages (see the Appendix for the data used).
(e) In the experiment Rounding vs. Decimals, we rounded the original numbers, always using modifiers when there was a loss of precision, *i.e. un poco ('a little'), casi ('almost').*
(f) In the experiment *Percentages vs. Fractions*, we selected frequently occurring percentages and their corresponding fractions

Comprehension Questionnaires. Each questionnaire is composed by two items, one inferential question *i.e.* to be answered correctly the question requires a deep understanding of the content of the text) and one question about details of the text which were expressed using a numerical expression. For each text there were two comprehension questions. The first was about the overall content of the text and the second was about the numerical part of the text. The later question always was phrased with

[4] The CYMK are crème (FAFAC8) and black (000000). Color difference: 700, Brightness difference: 244 [45].

the same number representation that was used in the text. Two examples, an inferential question (i) and a question devoted to details (ii) are given in Figure 1.

(i) El texto trata sobre:
('The text is about':)
  (a) La descomposición de una hamburguesa. *('The decomposition of a burger.')*
  (b) La creación de una hamburguesa. *('The creation of a burger.')*
  (c) La composición de una hamburguesa. *('The composition of a hamburger'.)*

(ii) Una porción de patatas fritas tiene un máximo de: *('One portion of french fries has a maximum of':)*
  (a) 200 kilocalorías. *('200 kilocalories.')*
  (b) 300 kilocalorías. *('300 kilocalories.')*
  (c) 400 kilocalorías. *('400 kilocalories'.)*

Fig. 1. Two examples of questions in the comprehension tests

Preference Questionnaires. Each participant performed a questionnaire with 20 questions that were rated using a 5 level Likert scale. For 10 of the statements the participant was asked about how easy was to read the text, that is, readability, while for the other 10 statements the participant was asked about how easy was to comprehend the text, that is, understandability. Each of the statements contained a numerical expression, using one of the following representations: rounded, unrounded, percentage and fractions, where 8 were written in words and 12 in digits. In the Appendix we present the set of numerical expressions used.

3.4 Equipment

The eye tracker used was the Tobii T50 [49] that has a 17-inch TFT monitor with a resolution of 1024x768 pixels. The eye tracker was calibrated for each participant and the light focus was always in the same position. The distance between the participant and the eye tracker was constant (approximately 60 cm. or 24 in.) and controlled by using a fixed chair.

3.5 Procedure

The sessions were conducted at Universitat Pompeu Fabra and they lasted from 20 to 30 minutes, depending on the participant. In each session, the participant was alone with the interviewer in the quiet room prepared for the study, performing the following three steps.

First, we began with a questionnaire designed to collect demographic information. Second, we conducted the experiments that were recorded using eye-tracking. Out of the 72 participants, 40 participated in the experiment *Digits vs. Words* (20 D, 20 N) and 32 carried out the *Rounding vs. Decimals* experiment (16 D, 16 N). Finally, The experiment *Percentages vs. Fractions* was carried out by 32 participants (16 D, 16 N) who had already participated in either the first or the second experiment, were not tired of reading, and were willing to read more. The participant was asked to read the texts in silence and to complete the comprehension tests. Third, we carried out a user preferences survey.

4 Results

In this section we present the analysis of the results of the eye-tracking and comprehension tests. The measures used for the comparison of the text passages were the means of the fixation duration, the total duration of reading, and the correct answers. In order to test our hypotheses, differences between groups and conditions were tested by means of (Bonferroni-corrected) Student's t-tests.

Shorter fixations are preferred to longer ones because faster reading is related to more readable texts [51]. We compare readability with understandability through the inferential items of the comprehension questionnaires.

First, we study the differences between both groups. Then, we analyze the impact of the different numerical expressions in readability and understandability.

4.1 Digits vs. Words

We did not find statistical significance in readability ($p < 0.444$) for group N (see Table 1). However, we found statistical significance for readability in group D taking into account the mean of fixation time ($p < 0.054$). This result supports our H1.1 hypothesis.

No statistical significance was found for both groups. Hence we reject hypothesis H1.2 ($p < 0.241$ for group N and $p < 0.269$ for group D).

4.2 Rounding vs. Decimals

We reject hypothesis H2.1, since we did not find statistical significance in readability in group N ($p < 0.867$) nor in group D ($p < 0.685$) when reading texts with rounded numerical expressions taking into account the mean of fixation time (see Table 2).

We also refute H2.2 because we did not found statistical significance for understandability in both groups ($p < 0.310$ in group N and $p < 0.695$ in group D).

Table 1. Experimental results for *Digits vs. Words*

Measure (ave. ± std.dev.)	[+Digit]	[−Digit]
	Group N	
Fixations Duration	0.19 ± 0.03	0.19 ± 0.02
Correct Answers	87.50%	95.00%
	Group D	
Fixations Duration	0.21 ± 0.04	0.24 ± 0.05
Correct Answers	85.35%	76.51%

Table 2. Experimental results for *Rounding vs. Decimals*

Measure (ave. $\pm$ std.dev.)	[+Round]	[−Round]
	Group N	
Fixations Duration	0.19 ± 0.03	0.19 ± 0.03
Correct Answers	97.22%	90.87%
	Group D	
Fixations Duration	0.23 ± 0.04	0.22 ± 0.04
Correct Answers	90.63%	91.67%

Table 3. Experimental results for *Percentages vs. Fractions*

Measure (ave. $\pm$ std.dev.)	[+Percentage]	[−Percentage]
	Group N	
Fixations Duration	0.19 ± 0.03	0.18 ± 0.03
Correct Answers	88.19%	96.86%
	Group D	
Fixations Duration	0.20 ± 0.06	0.23 ± 0.05
Correct Answers	80.55%	88.89%

4.3 Percentages vs. Fractions

We did not find statistical significance in readability for group N ($p < 0.462$) taking into account the mean of fixation time (see Table 3). However, our results confirm H3.1 because we found statistical significance for readability in group D ($p < 0.046$) when reading texts with numerical expressions in percentages. This group reads faster texts with expressions in percentages than texts with numerical information in fractions.

On the other hand, we reject H3.2 because we did not found statistical significance results for understandability in both groups ($p < 0.170$ for group N and $p < 0.474$ for group D, see again Table 3.

4.4 Survey

We found a high Pearson correlation of 0.95 between the answers of both groups. Therefore, groups N and D generally agreed in their answers with respect to readability and understandability. In Figure 2 we show the histograms of the survey results. Then, we calculated the standard deviation for all the statements. For the readability answers the standard deviation was significantly higher in group D ($\mu = 1.20$ seconds, $s = 0.41$) than in group N ($\mu = 0.87$ seconds, $s = 0.46$), with $p = 0.007$. This means that participants with dyslexia had a higher variability in their ratings. For the understandability answers, we found no significant difference between groups in their standard deviation ($p = 0.157$) for group D ($\mu = 1.04$ seconds, $s = 0.44$) and for group N ($\mu = 0.89$ seconds, $s = 0.37$).

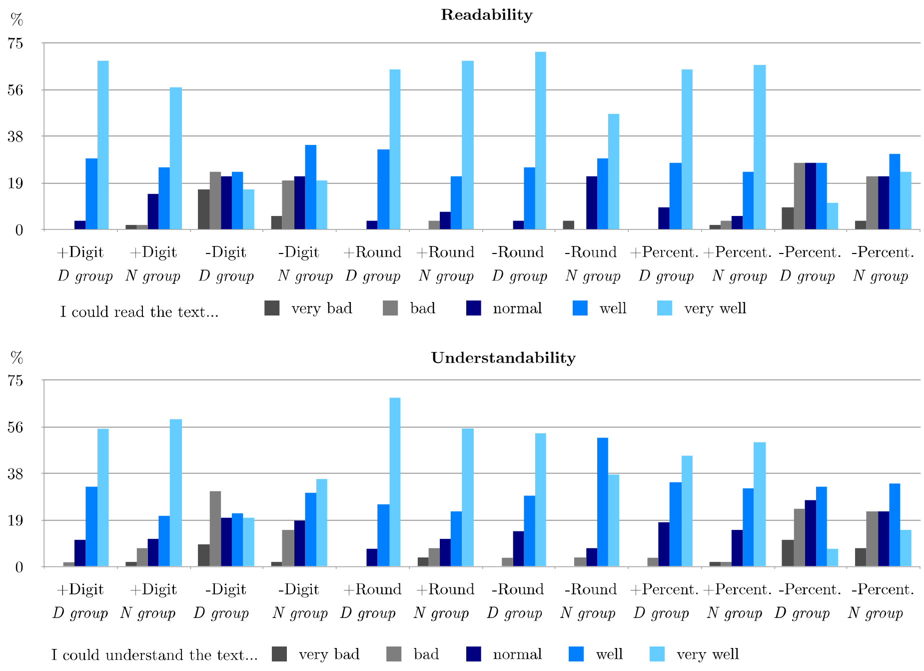


Fig. 2. Survey results for readability and understandability

Digits vs. Words. Participants significantly found numbers written in digits more readable than in letters ($p < 0.001$) as well as more understandable ($p < 0.001$). We also found significance within groups for readability ($p < 0.001$ in group D and $p < 0.001$ in group N) and understandability ($p < 0.001$ in group D and $p = 0.014$ in group N).

Rounding vs. Decimals. We did not find statistical significance for readability between rounded and unrounded numbers ($p = 0.272$) nor for understandability ($p = 0.446$). No significant differences were found within groups for readability ($t(53.937) = 0.479$, $p = 0.634$ in group D and $p = 0.111$ in group N) and understandability ($p = 0.163$ in group D and $p = 0.888$ in group N).

Percentages vs. Fractions. Participants significantly found percentages more readable than fractions ($p < 0.001$) as well as more understandable ($p < 0.001$). We also found significance within groups for readability ($p < 0.001$ in group D and $p < 0.001$ in group N) and understandability ($p < 0.001$ in group D and $p < 0.001$ in group N).

5 Discussion

With respect to differences between the use of digits and the use of numerical expressions in words, results indicate a statistically significant improvement in performance for readability in people with dyslexia when digits are employed. In contrast, in group N, we found no significant differences. This agrees with the fact that numerical expressions described using words require a longer number of words and/or characters in comparison with the corresponding versions using digits. Overall length is an already known parameter that creates difficulties for people with dyslexia, so the reduction in length involved in phrasing a number in digits should make it easier to read for them. Results for understandability in experiment *digits vs. words* are not statistically significant for both groups.

With respect to differences between the use of rounded and unrounded numbers, none of the differences found are statistically significant. Rounding numbers even with modifiers such as "around" or "almost" did not have the expected effect in our experiments.

With respect to differences between the use of percentages and fractions, there is a statistically significant increase in readability for group D when percentages are used instead of fractions. In contrast, no differences were found for group N. Again, results for understandability are not statistically significant in either case. There is an apparent contradiction in that for group D percentages seem to be easier to read but more difficult to understand. A possible explanation might be related to the nature of these expressions. From a conceptual point of view, both percentages and fractions convey the relative proportion between two quantities: the value of the percentage and 100 in the case of percentages, and the value of the numerator and the value of the denominator in fractions. However, the reference value in the case of percentages is implicit (or conveyed by the % sign). This implies that for fractions, two quantities have to be read, whereas only one needs to be read for percentages. This may account for the comparative ease for group D of reading percentages (only one quantity to read) vs. fractions (two different quantities to read). Participants that took this experiment did it because they were not tired and still willing to read more. In most cases participants with dyslexia were adults, and as such they have reading skills that are similar to adults without dyslexia. Note that fixation duration for these participants is shorter when reading percentages. More over, the percentages used in the texts were the most commonly used (see Appendix) and the more frequent the word, the shorter the eye fixation [25].

The standard deviations of the survey reveal that people with dyslexia made a greater difference between readability and understandability than people without dyslexia. For people without dyslexia, easier reading was correlated to text comprehension while participants with dyslexia dissociated these two elements, perhaps due do the nature of dyslexia, which affects reading but not comprehension of the language.

The higher variability of scores indicates that for people with dyslexia the representation of numbers has a much bigger impact on readability. However, the correlation of the answers of people with or without dyslexia is high. Hence, both groups generally agree in their rates with respect to readability and understandability.

The significant difference in the number representations of the survey for *Digits vs. Words* and *Percentages vs. Fractions*, are consistent with the quantitative data from the eye tracker where we found significant variations. Hence, the performance and the preferences of our participants with respect to these number representations are consistent.

6 Conclusions and Future Work

The presence of numerical information in a text impacts its readability. We have analyzed some of the different representations of numerical expressions in text to study their influence in the reading process of people with or without dyslexia. For each experiment our hypotheses have been tested and depending on the kind of numerical information for each group we have obtained different results.

The main contribution of this research is that numbers represented as digits instead of words as well as percentages instead of fractions improve readability of people with dyslexia.

These results may prove to be of value in providing empirical basis for the development or refinement of guidelines for the simplification of text. These guidelines exist in very general form [21] and they are currently employed as reference in a number of efforts to improve accessibility of text for groups of users with special needs. An empirical grounding that correlates particular expressions with particular user groups would be a very positive contribution.

This work can motivate work on text simplification for textual accessibility regarding numerical expressions [3, 4, 38]. Also, these findings can have an impact on current interactive systems for people with dyslexia that modify the text presentation but not its content such as *Claro ScreenRuler Suite* [10], *SeeWord* [22] or *IDEAL e-Book reader* [31]. We plan to integrate these findings in *DysWebxia 2.0* [44], a tool that modifies text content and layout for people with dyslexia.

Another field in which these results can be expected to have an impact is that of the assessment of readability. In general terms, computational models for predicting readability of texts are used, like FOG [23], Flesch, Flesch-Kincaid [20], and SMOG [35]. Current efforts along this line are considering a number of factors like average number of characters per word and average syllables per word to predict a readability score, but include no specific account for numerical expressions. Based on the results presented here, an effort might be made to expand the set of features used in the assessment of readability to include numerical expressions.

Future work includes the evaluation of numerical representation of other target groups and other measures of readability. We also need to evaluate other representations for specific numerical expressions, for example to represent time.

Acknowledgments. We thank Mari-Carmen Marcos for her assistance with the eye-tracker hardware, Martin Pielot for his help with statistics, and Jerónimo Bautista Velázquez for his help in distributing our experiments announcement among experts. The research of the first author was funded by the Spanish Ministry of Education and Science (Grant TIN2009-14560-C03-01) and the Catalonian FI scholarship program while the research of the second, fourth and fifth authors was funded by the Spanish Ministry of Education and Science (Grant TIN2009-14659-C03-01 Project), Universidad Complutense de Madrid and Banco Santander Central Hispano (GR58/08 Research Group Grant).

References

1. Al-Wabil, A., Zaphiris, P., Wilson, S.: Web navigation for individuals with dyslexia: An exploratory study. In: Stephanidis, C. (ed.) Universal Access in HCI, Part I, HCII 2007. LNCS, vol. 4554, pp. 593–602. Springer, Heidelberg (2007)
2. Bautista, S., Drndarevic, B., Hervás, R., Saggion, H., Gervás, P.: Análisis de la Simplificación de Expresiones Numéricas en Español mediante un Estudio Empírico (Analysis of the Simplification of Numerical Expressions in Spanish through an Empirical Study). Linguamatica 4(2), 27–41 (2012)
3. Bautista, S., Hervás, R., Gervás, P., Power, R., Williams, S.: How to Make Numerical Information Accessible: Experimental Identification of Simplification Strategies. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) INTERACT 2011, Part I. LNCS, vol. 6946, pp. 57–64. Springer, Heidelberg (2011)
4. Bautista, S., Hervás, R., Gervás, P., Power, R., Williams, S.: A system for the simplification of numerical expressions at different levels of understandability. In: NAACL Workshop Natural Language Processing for Improving Textual Accessibility, NLP4ITA (2013)
5. Bott, S., Rello, L., Drndarevic, B., Saggion, H.: Can Spanish be simpler? LexSiS: Lexical simplification for Spanish. In: Proc. Coling 2012, Mumbay, India (December 2012)
6. Brysbaert, M.: Arabic number reading: On the nature of the numerical scale and the origin of phonological recoding. Journal of Experimental Psychology: General 124(4), 434–452 (1995)
7. Butterworth, B.: Foundational numerical capacities and the origins of dyscalculia. Trends in Cognitive Sciences 14(12), 534–541 (2010)
8. Carrillo, M.S., Alegría, J., Miranda, P., Pérez, S.: Evaluación de la dislexia en la escuela primaria: Prevalencia en español (Evaluation of dyslexia in primary school: The prevalence in Spanish). Escritos de Psicología 4(2), 35–44 (2011)
9. Carroll, J., Minnen, G., Canning, Y., Devlin, S., Tait, J.: Practical simplification of English newspaper text to assist aphasic readers. In: AAAI 1998 Workshop on Integrating Artificial Intelligence and Assistive Technology, Wisconsin (1998)
10. Clarosoftware (2012), `http://www.clarosoftware.com/index.php?cPath=348`
11. Cohen, L., Dehaene, S., Verstichel, P.: Number words and number non-words: A case of deep dyslexia extending to Arabic numerals. Brain 117, 267–279 (1994)
12. Damian, M.: Asymmetries in the processing of Arabic digits and number words. Memory & Cognition 32, 164–171 (2004)
13. Dehaene, S.: Varieties of numerical abilities. Cognition 44, 1–42 (1992)
14. Drndarevic, B., Saggion, H.: Towards automatic lexical simplification in Spanish: an empirical study. In: Proceedings of the NAACL HLT 2012 Workshop Predicting and Improving Text Readability for Target Reader Populations, PITR 2012 (2012)
15. Drndarević, B., Štajner, S., Bott, S., Bautista, S., Saggion, H.: Automatic text simplification in spanish: A comparative evaluation of complementing modules. In: Gelbukh, A. (ed.) CICLing 2013, Part II. LNCS, vol. 7817, pp. 488–500. Springer, Heidelberg (2013)
16. Eden, G., Stein, J., Wood, H., Wood, F.: Differences in eye movements and reading problems in dyslexic and normal children. Vision Research 34(10), 1345–1358 (1994)
17. Ellis, A.: Reading, writing and dyslexia. Erlbaum, London (1984)
18. Fias, W.: Two routes for the processing of verbal numbers: Evidence from the snarc effect. Psychological Research 65, 250–259 (2001)

19. Fias, W., Brysbaert, M., Geypens, F., d'Ydewalle, G.: The importance of magnitude information in numerical processing: Evidence from the snarc effect. Mathematical Cognition 2, 95–110 (1996)
20. Flesch, R.: A new readability yardstick. Journal of Applied Psychology 32, 221–233 (1948)
21. Freyhoff, G., Hess, G., Kerr, L., Menzel, E., Tronbacke, B., Veken, K.V.D.: European guidelines for the production of easy-to-read information for people with learning disability. Technical Report ISLMH, European Union (1998)
22. Gregor, P., Dickinson, A., Macaffer, A., Andreasen, P.: Seeword: a personal word processing environment for dyslexic computer users. British Journal of Educational Technology 34(3), 341–355 (2003)
23. Gunning, R.: The technique of clear writing. McGraw-Hill (1952)
24. Herrera, A., Macizo, P.: Cómo leemos los números? (How we read numbers?). Ciencia Cognitiva 6(2), 44–47 (2012)
25. Hyönä, J., Olson, R.: Eye fixation patterns among dyslexic and normal readers: Effects of word length and word frequency. Journal of Experimental Psychology: Learning, Memory, and Cognition 21(6), 1430 (1995)
26. Interagency Commission on Learning Disabilities. Learning Disabilities: A Report to the U.S. Congress. Government Printing Office, Washington, DC, U.S. (1987)
27. International Dyslexia Association. Definition of dyslexia. Based in the initial definition of the Research Committee of the Orton Dyslexia Society, former name of the IDA, done in 1994 (2011), `http://interdys.org/DyslexiaDefinition.htm`
28. Ito, Y., Hatta, T.: Semantic processing of Arabic, Kanji and Kana numbers: Evidence from interference in physical and numerical size judgments. Memory & Cognition 31, 360–368 (2003)
29. Jiménez, J.E., Guzmán, R., Rodríguez, C., Artiles, C.: Prevalencia de las dificultades específicas de aprendizaje: La dislexia en español (the prevalence of specific learning difficulties: Dyslexia in Spanish). Anales de Psicología 25(1), 78–85 (2009)
30. Just, M., Carpenter, P.: A theory of reading: From eye fixations to comprehension. Psychological Review 87, 329–354 (1980)
31. Kanvinde, G., Rello, L., Baeza-Yates, R.: IDEAL: a dyslexic-friendly e-book reader (poster). In: Proc. ASSETS 2012, Boulder, USA. ACM Press (October 2012)
32. Landerl, K., Bevan, A., Butterworth, B., et al.: Developmental dyscalculia and basic numerical capacities: A study of 8-9-year-old students. Cognition 93(2), 99–125 (2004)
33. Landerl, K., Fussenegger, B., Moll, K., Willburger, E., et al.: Dyslexia and dyscalculia: two learning disorders with different cognitive profiles. Journal of Experimental Child Psychology 103(3), 309 (2009)
34. McCloskey, M., Caramazza, A., Basili, A.: Cognitive mechanisms in number processing and calculation: Evidence from dyscalculia. Brian and Cognition 4, 171–196 (1985)
35. McLaughlin, G.H.: Smog grading: A new readability formula. Journal of Reading 12(8), 639–646 (1969)
36. U. Nations. Standard Rules on the Equalization of Opportunities for Persons with Disabilities. Technical report, United Nations (1994)
37. Newell, A., Booth, L.: The use of lexical and spelling aids with dyslexics. Computers and Literacy, 35–44 (1991)
38. Power, R., Williams, S.: Generating numerical approximations. Computational Linguistics 38(1) (2012)
39. Rayner, K.: Eye movements in reading and information processing: 20 years of research. Psychological Bulletin 124, 372–422 (1998)

40. Rayner, K., Duffy, S.: Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. Memory & Cognition 14(3), 191–201 (1986)
41. Rello, L., Baeza-Yates, R.: Lexical quality as a proxy for web text understandability (poster). In: The 21st International World Wide Web Conference (WWW 2012), Lyon, France (April 2012)
42. Rello, L., Baeza-Yates, R.: The presence of English and Spanish dyslexia in the Web. New Review of Hypermedia and Multimedia 8, 131–158 (2012)
43. Rello, L., Baeza-Yates, R., Dempere-Marco, L., Saggion, H.: Frequent words improve readability and short words improve understandability for people with dyslexia. In: Kotzé, P., Marsden, G., Lindgaard, G., Wesson, J., Winckler, M. (eds.) INTERACT 2013, Part IV. LNCS, vol. 8120, pp. 203–219. Springer, Heidelberg (2013)
44. Rello, L., Baeza-Yates, R., Saggion, H., Bott, S., Carlini, R., Bayarri, C., Gòrriz, A., Gupta, S., Kanvinde, G., Topac, V.: Dyswebxia 2.0!: Accessible text for people with dyslexia (demo). In: Proc. W4A 2013, The Paciello Group Web Accessibility Challenge, Rio de Janeiro, Brazil (2013)
45. Rello, L., Kanvinde, G., Baeza-Yates, R.: Layout guidelines for web text and a web service to improve accessibility for dyslexics. In: International Cross Disciplinary Conference on Web Accessibility (W4A 2014), Lyon, France. ACM Press (April 2012)
46. Salguero, M., Alameda, J.: El procesamiento de los números y sus implicaciones educativas (Number processing and its educational implications). XXI Revista de Educación (Education Journal) 5, 181–189 (2003)
47. Sereno, S., Rayner, K.: Measuring word recognition in reading: eye movements and event-related potentials. Trends in Cognitive Sciences 7(11), 489–493 (2003)
48. Stanley, G., Smith, G., Howell, E.: Eye-movements and sequential tracking in dyslexic and control children. British Journal of Psychology 74(2), 181–187 (1983)
49. Tobii Technology. Product description Tobii 50 Series (2005)
50. Vellutino, F., Fletcher, J., Snowling, M., Scanlon, D.: Specific reading disability (dyslexia): What have we learned in the past four decades? Journal of Child Psychology and Psychiatry 45(1), 2–40 (2004)
51. Williams, S., Reiter, E., Osman, L.: Experiments with discourse-level choices and readability. In: Proceedings of the 9th European Workshop on Natural Language Generation (ENLG 2003), Budapest, Hungary (2003)

Appendix: Synonyms Pairs

Experiment 1: *Digits vs. Words* and **Experiment 2**: *Rounding vs. Decimals* (the set of target numerical expression used for each experiment are written in brackets, that is, [*Digits / Words - Decimals / Rounding*]).

Composición de una hamburguesa: El pan supone entre el [30% / treinta por ciento - 18,53% / casi el 20%] y el [50% / cincuenta por ciento - 29,57% / casi el 30%] del peso de una hamburguesa. La hamburguesa tiene un valor energético que oscila entre las [250 / doscientas cincuenta - 297 / casi 300] y [300 / trescientas - 398 / casi 400] kilocalorías. Un adulto con actividad moderada necesita en torno a [2.500 / dos mil quinientas - 2.489 / unas 2.500] kilocalorías diarias, por lo que una hamburguesa a la semana no desequilibra ninguna dieta ni siquiera incorporándole un sobre de [11 / once - 11,8 / casi 12] gramos de ketchup, que contiene [70 / setenta - 70,8 / un poco más de 70] kilocalorías.

Composition of a burger: The bread is between [30% / thirty percent - 18.53% / almost 20%] and [50% / fifty percent - 29.57% / almost 30%] of the weight of a hamburger. The burger has an energy value of between [250 / two hundred fifty - 297 / almost 300] and [300 / three hundred - 398 / almost 400] kilocalories. A moderately active adult needs about [2,500 / two thousand five hundred - 2,489 / around 2.500] kilocalories a day, so a burger a week not even unbalanced diet incorporating any one of [11 / eleven - 11.8 / almost 12] grams of ketchup, which contains [70 / seventy - 70.8 / a little more than 70] calories.

Composición de las patatas fritas: Las patatas fritas, a pesar de tener [3,6 / tres coma seis - 3,67 / casi 4] gramos de grasa y [234 / doscientas treinta y cuatro - 214 / un poco más de 200] kilocalorías por cada [100 / cien - 89 / casi 100] gramos, son un alimento muy energético por lo que ha de consumirse en pequeñas cantidades y esporádicamente. Contienen [11 / once - 11,82 / un poco más de 10] gramos de hidratos de carbono con un índice glucémico de [70 / setenta - 68,67 / casi 70]. Es decir, suponen un [50% / cincuenta por ciento - 58% / casi un 60%] del consumo diario recomendado de hidratos en mujeres y el [30% / treinta por ciento - 29,12% / casi un 30%] en hombres.

Composition of french fries: French fries, despite having [3.6 / three point six - 3.67 / almost 4] grams of fat and [234 / two hundred thirty-four - 214 / a little more than 200] kilocalories per [100 / hundred - 89 / almost 100] grams, are a very energetic food which has to be consumed in small quantities and sporadically. Containing [11 / eleven - 11.82 / a little more than 10] grams of carbohydrates with a glycemic index of [70 / seventy - 68.67 / almost 70] That is, suppose [50% / fifty percent - 58% / almost 60%] of the recommended daily intake of carbohydrates in women and [30% / thirty percent - 29.12% / almost 30%] in men.

Experiment 3: *Percentages vs. Fractions*

Composición de una hamburguesa: El pan supone entre el [25% / 1/4] y el [50% / 1/2] del peso de una hamburguesa incluyendo el [75% / 3/4] de los hidratos de carbono de esta. Estos hidratos suponen el [20% / 1/5] del consumo diario recomendado para un adulto con actividad moderada.

Composition of a burger: The bread is between [25% / 1/4] and [50% / 1/2] by weight of a hamburger including [75% / 3/4] of this carbohydrate. These hydrates represent [20% / 1/5] of the recommended daily intake for an adult with moderate activity.

Composición de las patatas fritas: Alrededor del [50% / 1/2] de los componentes de las patatas fritas son hidratos de carbono con un índice glucémico del [75% /

3/4]. Es decir, suponen un [25% / 1/4] del consumo diario recomendado de hidratos en mujeres y el [20% / 1/5] en hombres.

Composition of French fries: Approximately [50% / 1/2] of the components of French fries are carbohydrates with a low glycemic index of [75% / 3/4]. That is, suppose [25% / 1/4] of the recommended daily intake of carbohydrates in females and [20% 1/5] in males.

Numerical expressions used in the survey:

18.3%, menos de 1/5 ('less than 1/5'), menos del 20% ('less than 20%'), menos de un quinto ('less than a fifth'), menos de veinte por ciento ('less than twenty per cent'), más de 1/4 ('more than 1/4'), más del 25% ('more than 25%'), más de un cuarto ('more than a quarter'), más del veinte y cinco por ciento ('more than twenty-five per cent'), 27%, casi el 50% ('almost 50%'), menos de la mitad ('less than a half'), casi el cincuenta por ciento ('almost fifty per cent'), 48.6%, menos de 1/2 ('less than 1/2'), más de tres cuartos ('more than three quarters'), más del setenta y cinco por ciento ('more than seventy-five per cent'), 76.3%, más de 3/4 ('more than 3/4') and más del 75% ('more than 75%').

Studying a Head Tracking Technique for First-Person-Shooter Games in a Home Setting

Torben Sko[1], Henry Gardner[1], and Michael Martin[2]

[1] The Research School of Computer Science, The Australian National University
[2] The Research School of Finance, Actuarial Studies and Applied Statistics, The Australian National University
{torben.sko,henry.gardner,michael.martin}@anu.edu.au

Abstract. This paper examines webcam-enabled head tracking for games in a home setting. A new head interaction technique was developed based upon prior laboratory-based research, with a focus on making it robust to the variable conditions of a home setting. Our technique was integrated into a test-bed game and 550 hours of gameplay data was collected from 2500 users, many of whom also provided formal feedback. The head tracking performed creditably and players reported that the experience was more immersive. Head tracking failed to enhance competitive playing performance, perhaps owing to familiarization effects. Nevertheless, the data revealed evidence of learning amongst users, suggesting that performance would improve with continued use. Key lessons that emerged in the home setting in contrast to the earlier laboratory study were a demonstrated need for clear guidance and feedback during system set-up, and greater caution regarding its deployment, having discovered a small population of users who became nauseous.

Keywords: head tracking, gestural interaction, online studies, games.

1 Introduction

In recent years, webcams have become a ubiquitous form of technology. When placed in a desktop setting – where a user sits up-close to their computer screens – webcams can be used to track the 3D location and orientation of the user's head position. Despite the availability of this technology, head tracking has seen relatively little adoption, even in computer games that have elsewhere seen a significant move towards gestural interaction. Our previous head tracking work [23] highlighted several barriers to adoption, with the most prominent of these being the perceived reliability of the tracking system. This issue was presented in the context of a laboratory setting and techniques of a prototype quality. Consequently, it could only be hypothesized what the real-world impact of this issue was – when considered in the context of a home setting and a consumer-grade technique.

This paper presents the first study of its kind to examine a refined head tracking technique in a household setting. We conducted a remote user study with a built-in logging system to allow us observe how our technique functioned when used by home

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 246–263, 2013.

players. In doing so, we looked to examine whether the technique could work reliably enough within this uncontrolled environment and provide enough benefit that players would adopt it. In examining this adoption, we were also keen to explore what impact our head-tracked interaction technique had on the user's competitive playing performance, as compared to a button alternative. Such an investigation responds to the idea that "hardcore" PC players can be quite dismissive of gestural interfaces on the basis that they do not yield any playing performance benefits [11]. The public exposure the study entailed also provided an opportunity to examine how a large audience of players reacted to the idea of using head tracking in games – a concept that still remains quite novel. This paper examines this reaction in an effort to identify what qualities may help to make a head tracking technique more appealing.

The next section provides some background and further motivates the decision to conduct the study in a home setting. Sections 3 and 4 present a refined interaction technique and detail the study used to assess it. Section 5 presents the results of the study, while Section 6 summarizes our findings as a set of head-tracking guidelines, designed to help others develop their own interaction techniques.

2 Background

Head tracking was first reported in the literature in the late 1960's by Sutherland [26] who attached a mechanical arm to a user's head to detect their head pose. By contrast, modern head-tracking libraries, such as [22, 28], can now function with just a standard webcam (see [17] for a full head tracking review). Following these improvements, researchers have explored the use of head tracking within various desktop applications, including gaming [3, 12]. For example, the work of Wang et al. [27] presented a head-based leaning technique in a first-person-shooter (FPS) game that allowed players to dodge bullets. Their work found that the inclusion of head tracking improved a measure of player *presence*. The commercial game, *ArmA 2* [2], adopted a similar technique that also allowed players to look around independently of their aim. This title required use of the infrared-based hardware system, TrackIR [18]. Another tracking system more recently released, is the Kinect [16]. Despite the availability of this hardware, only a few games exist that specifically utilize the head as a gestural controller.

In the absence of any widespread adoption, our previous work [23] examined the ways in which head tracking could be utilized in an FPS game, by comparing Wang's leaning technique [27] to six other head-interaction techniques. In this work we implemented and evaluated our techniques using focus groups. The general reception was positive, although gamers were also concerned about the *reliability* and *robustness* of the tracking system. As this work was conducted in a controlled, laboratory setting, how these concerns might translate to real-world use within a household setting was a matter for conjecture.

Fig. 1. A user performing a right lean. The amount the player needs to tilt is kept very subtle to ensure the interaction technique is not too strenuous. During the lean the screen and gun model are both rotated and the player shifts sideways.

Many existing studies have acknowledged the importance of the real-world environment in assessing the usability of new technology. For example, Hartson [13] argues that both the test settings, along with properties of the system itself, are intrinsic factors that impact upon the usage patterns, making it difficult to replicate in a laboratory setting. The work of Brown [5] similarly points out (in the context of a geo-locating tool) that while the lab provides a good setting for exploring the technical feasibility, it provides little insight into *user acceptance*. These issues, that of the real-world realization and adoption, are particularly relevant when discussing webcam-based head tracking given the use of vision-based algorithms. The performance of the tracking is susceptible to environmental factors, such as lighting conditions and physical set up – making it necessary to use a home setting if we wish to know its real-world performance and adoption.

To help deliver studies within remote settings, researchers have increasingly taken to using the Internet. The benefits of online studies include the ability to reach potentially thousands of home users [7], across a large and varied sample [29]. For example, Kittur et al. demonstrated that is possible to utilize existing crowdsourcing systems (i.e. systems where users complete small tasks for monetary incentives) for the purposes of running a user study [14]. The level of sophistication employed varies dramatically, from simple surveys [10] to sophisticated remote studies where the participant and researcher are separated *physically* and *temporally* [13]. In these latter studies, a logging system is employed to capture how the user interacts with the study software, which frees the user to engage with the study in their own time [21]. For example, the work of Costanza et al. [8] used a logging system to capture the visual and auditory output of their software, allowing them to later replay its use. Some of the other feedback mechanisms that researchers have used include; text and video chat, questionnaires, interviews, forums, and diaries [1]. Our own study employed many of these concepts, as described in Section 4.

3 Technique Description

The head interaction technique presented in this paper is referred to as the Handy-cam and Leaning (HAL) technique (pictured in Fig. 1). It combines and improves upon the two most promising techniques identified in our earlier work [23]. The first of these was a *leaning* technique, which was first explored by Wang et al. [27]. This technique allowed a player to perform an in-game lean movement by physically leaning (that is, tilting and/or offsetting their own head). The resulting in-game action allows players peer out from behind cover so they can spy on enemies and better protect themselves during a gunfight. When featured in an FPS title, this form of control is typically mapped to the ‘Q’ and ‘E’ keys. By comparison, using head-movements to perform this form of control is potentially more intuitive [23] and can provide a finer level of control through its continuous nature [25]. In Section 5.2, we examine whether this increased level of control can indeed yield a stronger playing performance.

The second technique chosen was the *handy-cam* effect, which aimed to reinforce the intention that a first-person game is being viewed through the eyes of the protagonist. The technique worked by adding a “wobble" to the virtual camera so that the player’s own subtle and involuntary head movements were mimicked by the virtual camera. For example, if the user was pitching their head slightly upwards, the virtual camera was subsequently pitched slightly upwards (likewise for the other head axes). A similar effect is used in modern cinema in films like “The Bourne Identity” (2002), where a subtle wobble is added to the view to make it look as if the action is shot on a handheld camera. This effect can make the viewer feel more *present*, as if they are viewing the action directly with their own eyes [15]. Adding a jitter has also been found to make the scene appear more *realistic*, by increasing the perception of self-motion [6]. Several studies have examined the effects that a *pre-animated* jitter can provide, both in pre-recorded video [20] and games [9]. Our earlier work was the first known attempt to create an interactive version of this effect, where it was unanimously well received by the focus group participants [23].

3.1 Technique Implementation

The design of HAL has seen our earlier prototype techniques [23] greatly matured into a technique designed for mass-use. More specifically, whereas our previous implementations were targeted at a single computer set up, HAL needed to work across a multitude of home computer configurations. To make the technique cope with variations in the set up, the following enhancements were added:

- No assumptions were made about the user’s resting position in order to cater for differences in the physical configuration of the home systems.
- Time-based smoothing (rather than sample-size based) was used to account for computer speed variability.
- Smoothing of the head data was adapted based on the tracking quality, thereby allowing HAL to remain responsive on slower systems.
- An easing curve was added to the leaning for finer control over smaller lean movements.

- Several minor tweaks were added to avoid any negative visual impacts the head technique may create. These included constraining the range of head mapped movements and fadding out/in following a head loss/aqausition.

Our HAL technique can be thought of as a mapping from the user's head position to a normalized *leaning-amount* and a *camera offset*. The head data was obtained using Seeing Machines' faceAPI (v4) software [22], which provides both the position and orientation of the user's head. From this data, the leaning-amount was computed using both the roll and sideways head movement (as both axes are typically involved in a leaning action). To derive the leaning-amount the following logic was applied:

```
01. rollResting = UpdateMean(rollResting, headData[ROLL], rollRange);
02. roll = headData[ROLL] - rollResting;
03. roll = Normalize(roll, rollMin, rollMin + rollRange * leanScale);
    ... // Repeat the above for sideways movements
04. lean = Normalize(roll + sidew, 0, 1);
05. lean = lean * (1 - easeAmount) + Ease(lean) * easeAmount;
06. lean = MovingAverage(lean, leanSampleTimeframe);
07. lean = FadeIn(lean, 2);
```

The bolded variables could be tuned using the settings panel, thereby allowing users to customize how the technique functioned. In lines 01 and 02, the head value is converted from an absolute position into an offset relative to the user's resting position, allowing us to identify how much they have moved. As the head position is stated relative to the webcam, the user's resting position will be a non-zero position vector (e.g. if the user sits 80cm away, the reported depth will be 0.8). There are several sources of variation that will influence this position, including the camera's position, the user's seating position, as well as how the software interprets the user's face. In our earlier work [23] this position was assumed, which caused the technique to behave incorrectly for some participants. Our HAL technique instead employed a running mean to compute this resting position. To avoid having this position shift too much during play, the mean was only updated when the new head data was within a certain range of the mean (*rollRange* = 30°, *sidewRange* = 15cm).

In line 03, the roll amount is normalized over a usage range. This range is scaled (*leanScale* = 1, *sidewScale* = 1) in order to control the sensitivity of the technique. A minimum is used (*rollMin* = 3°, *sidewMin* = 2cm) to introduce a “dead zone" or no-movement zone. This dead zone ensures subtle head movements avoid invoking a lean movement, which could otherwise interfere with the player's aim. In line 04, the two components of the lean movement are combined and normalized between 1 (ful left lean) and -1 (full right lean). In line 05, an ease-in/ease-out curve is applied to the lean-value based on the *easeAmount* (= 100%). The leaning value is then smoothed using a moving average (line 06, *leanSampleTimeframe* = 0.2sec) and faded (line 07) in the event that the user has only started to be tracked within the last two seconds.

The lean-amount obtained from the previous logic was then mapped to a leaning action. In the case of our game, this corresponded to a sideways player movement, a gun movement, and a slight reduction in the field of view (FOV). The gun movement and FOV reduction together recreated some of the look-and-feel of the *iron sighting* technique, which was well received in the previous focus group studies [23]).

The other key aspect of HAL was the handy-cam effect. To compute the camera offset, the following logic was applied:

```
01.   leanMean = MovingAverage(lean, 0.5);
02.   pResting = UpdateMean(pResting, headData[PITCH]);
03.   p = headData[PITCH] - pResting;
04.   p = p * handycamScale * pitchScale;
05.   p = p * (1 - (leanMean * dampenWhileLeaning));
06.   p = MovingAverage(p, handycamSampleTimeframe);
07.   p = FadeIn(p, 2);
08.   handycamOffset[PITCH] = p;
      ... // Repeat the above for all axes except head depth
09.   cameraPosition += handycamOffset;
```

As before, the head data is converted into an offset (lines 02 and 03, above). In line 04, the effect is scaled using both a global scale (*handycamScale* = 1) and an individual scale for each axis (*pitchScale* = 1, *rollScale* = 1, etc). These scaling factors allowed the user to control the strength of the technique. In line 05, the offset is potentially dampened during times that the user was leaning (*dampenWhileLeaning* = 50%). This was included to avoid having the effect interfere with the player's aim whilst leaning. The value was then smoothed (line 06, *handycamSampleTimeframe* = 0.2sec) and faded (line 07). This was repeated for all the other head axes, except for movements along the forward-facing axis as these were found to negatively impact on the appearance of the weapon model positioned in the foreground of the scene. The computed offset vector was added to camera's position (line 09) to create the handy-cam effect. A complete implementation of the HAL technique can be found online[1]. It has been implemented using freely available software.

4 Study Design

To study HAL we wanted to observe; how it performed, to what degree it was adopted by players, and what players thought of it. To achieve this we implemented HAL into the commercial FPS title, Half-Life 2 (HL2). This allowed us to utilize a wealth of existing game content, which provided a rich setting in which to examine our technique. As HL2 does not include leaning as part of its control set, we modified the game to include it. We also constructed a custom lean-centric game mode to explore the hypothesis that a greater focus on head-control would elicit a more positive user response [23]. This custom mode was called FaceOff Paintball (FOP). The following summarizes the differences in these game modes:

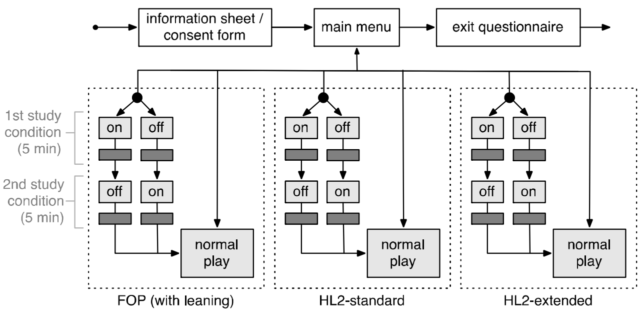
---

[1] `https://github.com/torbensko/HAL`, *accessed Apr '13*

- *HL2 mode:* existing game levels of Half-Life 2. The levels were not designed with leaning in mind, but can be played with leaning (controlled using either the keyboard or head movements) with our implementation.
- *FOP mode:* a novel paintball game level where players try to survive for as long as possible. The player (and enemies) only take one shot to be "marked" (i.e. eliminated) making it important to use the provided cover, which we felt could be best done through the use of leaning.

We added several mechanisms to both game modes that allowed us to examine how the HAL technique was performing and being used. Firstly, we monitored and logged characteristics of participants' use of the system including their head position, the leaning value, their playing performance (e.g. game score), system performance (e.g. frame-rate), the hardware (e.g. camera model), and potentially a picture of their set up as taken with their webcam. These data were only sent if the user opted into the study aspect of the game. At the time of downloading the game we surveyed the participants (via a sign-up form), asking them about their *age*, *gender*, *gaming experience* and their *enthusiasm towards head tracking*. We were able to associate this information to the logged data using a sign-in panel within the game.

To collect the player's opinion of the technique, the game was modified to include a post-play questionnaire, made up of 45 questions, most of which requested responses on a five-point Likert scale. We also incorporated an optional comparison test, which provided us a more formalized way of examining the impact head tracking had on the game experience. Within the test, users played the game twice, once with the *head tracking enabled*, and once with it *disabled* (randomized order between users). This test could be played under either HL2 or FOP game modes. Under the HL2 mode, players could choose to play with either the traditional control set (without



Game Mode	Head tracking: Enabled	Head tracking: Disabled
FOP (with leaning)	HAL	Buttons (Q and E to lean)
HL2-standard (no leaning)	Handy-cam only	Buttons
HL2-extended (with leaning)	HAL	Buttons (Q and E to lean)

Fig. 2. A flow diagram for the online experiment (left) and resulting comparison conditions (right). In the flow diagram, the mid-gray boxes represent states where the user is playing. The black dots indicate where the system auto-assigned users to a condition. The 'on' boxes denote the condition where head tracking was enabled, while in 'off' head tracking was disabled. The darkest boxes represent the in-game survey. The system forced the surveys to be viewed but did not prevent users dismissing it without answering any questions. The player could freely switch between any of the game modes and could complete the comparison test multiple times.

leaning) or with the extended control set (including leaning). This meant there were two conditions: head control (*enabled / disabled*) and game mode (*FOP / HL2-standard / HL2-extended*). When the playing under the HL2-standard mode (without leaning), only the handy-cam aspect of the head-interaction technique was in effect, allowing us to study its impact on its own. At the end of each condition, an in-game survey was presented to the users. A summary of the study is provided in Figure 2.

Our study system was placed online and promoted through the use of two YouTube videos. The release of the work generated a considerable amount of online discussion, which was found to provide a valuable source of feedback [21]. When combined with the open text response taken from the exit questionnaire (479 comments), a collection of over 2000 comments was accumulated. These comments reflected how HAL was both *perceived* (prior to using) and *received* (after using it).

5 Results

Playing data was collected from 2593 users from across 100 countries, with the US, Russia and the UK being the most common countries of origin. The most represented group was young males with “lots” of FPS experience. Through the use of a pre-play survey, administered during the sign-up process, we found that 75% professed to being “excited” about the use of head tracking. The median playing time was 7 minutes. A total of 550 hours of playing data was logged.

The remaining part of this section is split based on the main research goals; Sec. 5.1 examines how the technique performed in an uncontrolled environment, Sec. 5.2 explores how players adopted the technique and its impact on their playing performance, while Sec. 5.3 examines the players’ opinion of HAL as a way of identifying the qualities that have most influenced its reception.

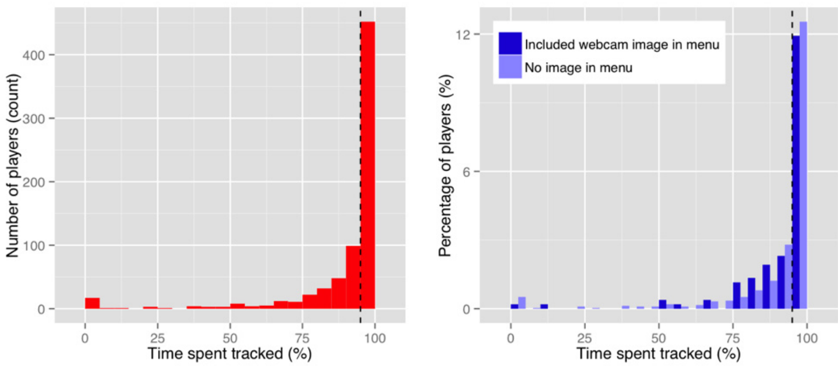


Fig. 3. System reliability: the players have been binned based on the percentage of time they were successfully tracked during the *tracking enabled* condition of the comparison test. Two versions of the study software were distributed, with one including an image of the user featured on the main menu – in the hope it would help them recognize deficiencies in their set up. No improvement can be observed when this image was included (as supported by a statistical comparison).

5.1 Technique Performance

By logging HAL's functional performance, we were able to examine how well it performed in a household setting. Figure 3 illustrates the reliability of the system by binning participants based on the tracking rates during the *tracking enabled* condition of the comparison test. Of the 726 participants that played under this mode, 62% (452) were tracked for 95% of their playing session or more. By contrast, 2% (17) were never tracked.

When examining the *complete* participant sample, it was noted that 23% (595) were *never* successfully tracked. Examining the hardware details of these players revealed that 34% (202) of these non-tracked players did not have a working camera – meaning that their camera was incompatible, incorrectly configured, or non-existent. The collected forum comments highlighted that built-in cameras were the most troublesome. The other 66% of the non-tracked players (393 users) did a have recognized camera, but were still never successfully tracked. Several images were collected from the user's own webcams to help explain why the tracking failed. Figure 4 demonstrates that some players used the system under far-from-ideal conditions. Some of the common issues encountered included: partially cropped and obscured faces (e.g. some users wore obscuring apparel, others sat too close); dimly lit settings (e.g. some users were lit via their monitor only); high contrast images (e.g. over-exposure of the image due to bright lighting); blurry/unclear images; and pre-processed images (e.g. with some users featuring partial overlays and virtual avatars). Because many of these issues could have been corrected on the user's side, the software was modified during the study to display an image from the user's own webcam on the main menu of the system in the hope that users would identify and rectify deficiencies in their own setups. As Fig. 3 shows, this had limited impact on tracking reliability, an observation verified by performing a classification of user images based on the apparent issues.

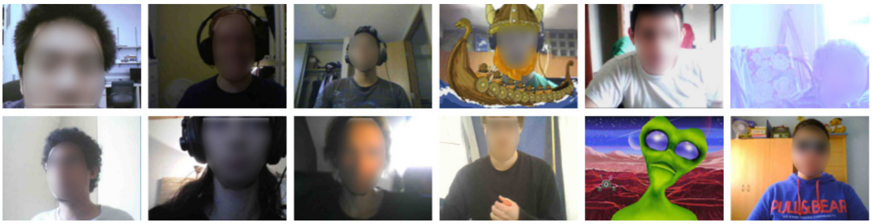


Fig. 4. Some of the webcam images (with faces blurred to protect identities). Many of these images show conditions that would make a user's head difficult to track: dim lighting, overexposure, obscured facial features (e.g. sun-glasses), a highly angled camera. In the case of the alien head and Viking boat, the webcam feed has been pre-processed before being received by the head-tracking library. Explicit permission was sought from users before collecting these images. The study was approved under the ethics protocol, 2010/032.

5.2 Player Adoption

The log data from the successfully tracked players provided an insight into how HAL was used by players. The data illustrated that the FOP game mode was successful in encouraging more leaning, with the head leaning being more popular than button leaning under both game modes. As one participant noted, *"I rarely make efficient use of the lean function in games because it usually requires some finger gymnastics to strafe and lean at the same time. I found myself leaning much more when all I had to do was move my head. I'm very pleased with this idea."* Button-based leaning saw almost no use within the HL2 game mode. As the game content did not encourage leaning, this outcome would suggest that the novelty of the leaning control alone was not enough to encourage players to use it. However, when the leaning control was performed via head tracking its use increased and was well sustained across the users who chose to play for more than an hour under the HL2 mode (30 players).

The playing logs also revealed how much players tended to move while using the technique. Figure 5 shows the amount of head movement under three different control schemes: HAL (both head leaning and handy-cam enabled), handy-cam only and head-controls disabled. In the disabled condition the player's head position was still logged despite having no apparent effect on the game, allowing us to observe the natural head movements that occur during play and establish a control group. The data in Fig. 5 show that players moved their head the least when the head control was disabled. Users appeared to move least in the vertical axis and the roll axis. By comparison, the pitch and yaw axes exhibited a greater amount of movement, making it more likely that they would yield false-positives if used for head controls. This result affirms our decision to develop a technique that predominantly relied on the roll axis.

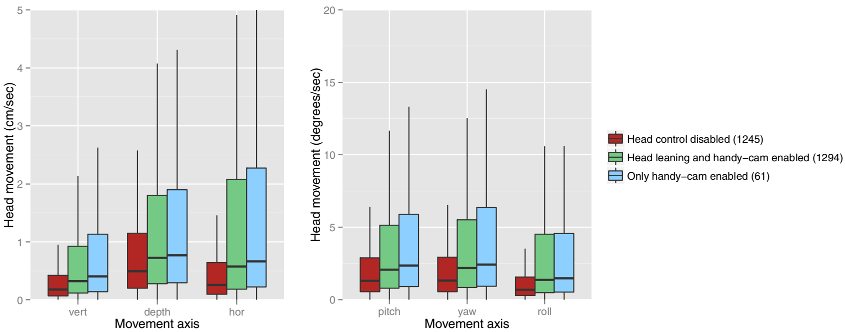


Fig. 5. The amount of head movement under different control schemes. The data has been sourced from the HL2 game mode during the comparison test. A half-second smoothing window has been used to account for tracking noise. The control schemes were taken from the different test conditions (see Fig. 2 for a list of conditions). The numbers in parenthesis indicate the number of applicable players, noting that only a few players chose to disable the leaning control, hence why the *handy-cam only* category is less populated. Nonparametric Wilcoxon tests reveal the *disabled/HAL* and *disabled/handy-cam* conditions to be significantly different ($p < 0.01$). By contrast, the *HAL/handy-cam* were not found to be significantly different.

The data shows that the amount of roll movement increased when the head leaning was enabled, however, even under this condition, this axis still exhibited less energy than the other two rotational axes.

Unexpectedly, the charts reveal that players tended to move most when only the handy-cam effect was enabled. It was expected that the HAL condition would exhibit the most movement, as part of performing the leaning control. By comparison, the handy-cam was intended to utilize the user's subtle and involuntary head movements. This finding suggests that users were trying to pro-actively use the handy-cam effect, presumably to look around, illustrating that its design intention was not well conveyed. When used in conjunction with the leaning, the amount of movement dropped. This may suggest that users were not as inclined to use the handy-cam for practical purposes if a more obvious use of head tracking was also present. For this reason, the handy-cam is probably best used alongside another head-interaction technique.

Returning to the topic of HAL's adoption, another aspect to consider was the impact that head tracking had on the users' playing performance. Past research shows that an interface that provides performance benefits is more appealing to players [11]. To explore HAL's performance impact, our system included an online leaderboard within the FOP mode, which globally ranked players based on their score. By including this indirect competition, it was hoped players would focus on their playing performance. A total of 24,000 attempts were recorded, lasting an average of 36 seconds. It was interesting to note that the top five players consisted three who chose not to lean, one player who mainly used button leaning (3^{rd} place) and one head-leaning player (4^{th}). Three of the top 10 players mainly used head leaning. A mixed-effects model was used to examine how button and head leaning affected the score.

For each attempt the number of marks (i.e. eliminations) was divided by the attempt duration to yield a *marks-per-second score*. The one-shot-eliminates nature of the FOP mode meant that the scores were not normally distributed, so a log of the marks-per-second score was taken (after removing scores of zero). The resulting "*log-score"* performance measure was normally distributed. We also added to the model aspects like *prior use* of the system (e.g. prior duration in FOP mode, prior head/button leaning), information about the attempt (e.g. duration of play, percentage of time tracked), as well as demographic and hardware information. The usernames were fit as a random effect in the model as observations from the same user were

Table 1. The coefficients of the mixed-effects model. The percentage of time spent head leaning includes the periods where the user performed only a partial lean movement.

	(Intercept)	Attempt duration	Time tracked (%)	Head leaning (%)	Button leaning (%)	Computer speed (frame-rate)	FPS experience: "lots"	Sex: "male"
Estimate	-2.6673	0.004	0.3273	-0.2456	0.5304	0.0008	0.0479	0.1627
Std. Err	0.0942	0.0002	0.0176	0.0166	0.0255	0.0001	0.0216	0.0918

likely to be correlated with one another, while observations between users were assumed independent. All other covariates were fit as fixed effects.

After an initial 'full' model fit, all clearly non-significant variables were excluded and reduced model was assessed using both deviance-based criteria and AIC (Akaike's Information Criterion). This reduction process was continued, until a final model was reached for which the remaining model terms were significant. The estimated coefficients, with standard errors, for the final model are shown in Table 1. The model indicates that gender (male) had a positive influence on log-score, as did having extensive FPS experience, a faster computer, as well as being reliably tracked. Our mixed-effects modeling revealed that button leaning had a positive influence on the leaderboard score, but that head-tracked leaning *negatively impacted* it. With users only playing for an average of seven minutes, it is possible that this detrimental effect was the result of a familiarization period. A correlation test found the accumulated duration of play had a significant positive correlation with the score (0.147), illustrating that a learning effect was present. This learning effect was also observed when examining the two forms of leaning. The accumulated duration spent button leaning and head leaning were both found to have a significant positive correlation with the score (0.123 and 0.107 respectively), albeit with the head leaning being harder to learn. This learning effect is encouraging, as it suggests that HAL may be able to eventually facilitate a better performance over time.

To supplement this test, we examined whether the participants' use of the leaning technique changed over time. Little change was observed over a one hour period, suggesting it probably takes several hours of play to become reasonably skilled using HAL. Such an outcome suggests a longitudinal study would be required to properly examine the playing performance impacts of head tracking. Such a study of competitive performance may benefit from adopting a richer control set than that used by us in the FOP mode in order to mitigate the trade-off that occurs between the *physical affordance* of an input method, the perceived *in-game benefit* and the *reliability* of the head tracking system. As one participant noted, *"The best keys for leaning are "Q" and "E", but I would rather use those keys for more important actions. If it works right this would free up some valuable real estate."* Similarly, having players compete directly against each other in the gameplay may help to address the familiarization performance slump. As one participant noted, *"It would be interesting to see if the "I'll stick to my keys" players would change their tune if they started getting owned by head tracking player because of this difference."*

5.3 Player Opinion

With head tracking remaining a relative novel concept within gaming, we wanted to discover what player's thought of HAL so that we may better understand what influences its appeal. The players' opinions were, in part, collected via the comparison test surveys and the exit questionnaire. An exploratory examination of the data was conducted using a classification tree (CART) approach [4]. This approach seeks to predict (classify) response variables based on the covariate values by recursively partitioning the covariate space through a sequence of binary splits on individual covariates. Each step in the algorithm chooses a covariate that provides the most homogeneous groups, delivering a tree-like structure for the model. The tree model approach does not require the pre-specification of interaction terms; it allows for complex multi-way interactions; and it caters for a mix of continuous and categorical variables without imposing any particular parametric structure, making it well suited to exploring rich datasets, such as those yielded by online studies [24].

Some of the covariates included within our tree modeling included: the *tracking reliability* and the *stated prior enthusiasm for head tracking* (to help explore the presence of an enthusiasm bias), *hardware performance* (e.g. frame-rate, to help explore whether a minimum hardware requirement exists), and the *prior playing time* (to test whether longer exposure improved the player's opinion). For the comparison test questions, the model also included the *game mode* (to test whether a greater focus on the head-technique yielded a more positive result), and the *head tracking condition* (i.e. enabled/disabled).

The CART trees for the exit questionnaire and in-game surveys both told a very consistent story. From the exit questionnaire responses, participants were clearly in favor of head tracking. For example, when asked, *"What are your thoughts about having the head tracking technique demonstrated in this game incorporated into other first-person-shooters?"* the resulting tree, comprising of 899 responses (85% in a positive category), first split on *prior enthusiasm*, with both first-level nodes classified as a positive response. A similar result was found when asking the more contentious question: *"Would you be inclined to use this head tracking technique in more demanding situations, such as when playing on harder difficulties or in competitive online games?"* Based on the 825 responses (55% in a positive category), again the tree split on *prior enthusiasm*. Encouragingly, when examining the response of just the "skeptical" or "undecided" players (i.e. the *non-excited* players), 65% (72) were still in favor of having HAL included in other games, however, only 26% (27) would use it in a competitive scenario.

For the comparison-test, the *head tracking* condition (disabled/enabled) was consistently the most dominant factor (primary splitting variable, or best-discriminating covariate) for each of the 18 questions. For example, one of the questions asked players how much they "enjoyed" the game. It received a total of 1276 responses from 888 users (includes responses to the different study conditions and under the different game modes). The resulting tree can in Fig. 6. The tree demonstrates that the *tracking condition* is the primary predictor, with the *prior enthusiasm* being a secondary effect relevant for users with tracking enabled, with no improvement in classification for users for whom tracking was disabled.

Although these results show that players liked head tracking, it seems that a prior enthusiasm bias was present. This phenomenon was probably a result of the recruitment approach, which was based on drumming up interest in the work by posting videos on the Internet. With nodes at lower levels of the tree exhibiting high misclassification error rates, this enthusiasm effect appears to have swamped other possible interactions within the data, thus limiting the ability of the model to reveal finer structure. In other words, the preponderance of enthusiastic participants in the study meant that the covariate space was insufficiently heterogeneous for variation within the other covariates to be able to adequately explain variation in the response after the enthusiasm effect was accounted for.

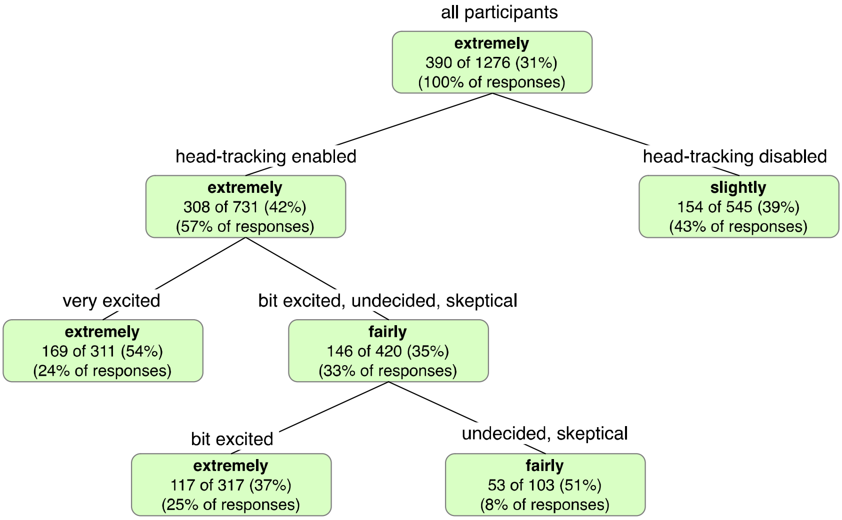


Fig. 6. Fitted decision tree model: users were asked to state how much they *enjoyed* the game experience using a five-point Likert scale ("extremely", "fairly", "moderately", "slightly", "not at all"). The tree initially splits on tracking condition (enabled/disabled), with users for whom tracking was disabled classified as enjoying it only "slightly". For users for whom tracking was enabled, predicted enjoyment depended on their reported prior enthusiasm for head tracking (categorized as "very excited" to "very skeptical"), with "excited" tracked users predicted to "extremely" enjoy it, and non-excited tracked users predicted to only "fairly" enjoy it. The details recorded at each node are the predicted classification, the proportion of data at the node consistent with that classification (x out of y, the classification at the node being the result of a majority vote) from which a classification rate can be calculated, and the proportion of the entire data at that node. The terminal nodes represent the final fit.

To further explore what people thought of HAL, we examined the written feedback collected via both the exit questionnaire and from the online discourse following the study. The comments helped identify the head tracking *benefits* that appeared to most appeal to players. Players frequently mentioned its ability to improve the game's *immersion* and *realism*. As one player stated, *"Even if it doesn't really change gameplay all that much, it certainly enhances the immersiveness of the 3D game environment, more so than higher frames per second, or higher resolution, or better shaders could hope to do at this point."* Several comments attributed these qualities to the *handy-cam* effect, feeling that the technique may be suited to horror style games, with the title *Amnesia* being repeatedly mentioned. The leaning aspect of the technique was also noted as being very *natural*; *"People have been doing this forever, leaning to the side in a game, hoping for a "better view", but this time your natural movement leads to an in-game result."*

There was strong evidence that head tracking appealed to a spectrum of PC players, including those initially skeptical about gestural control. One article about the work opined: *"Afraid that new camera-based interfaces herald the death of traditional gameplay? Fear not! This video is whispering in your ear, Core Gamer, and* *saying, 'Don't be afraid. Everything is going to be alright.'"*[2] In another example, the popular YouTube user, Robbaz: The King of Sweden, placed a self-narrated video of himself playing the study software onto YouTube. As dedicated FPS player – as evidenced by his 300 odd videos, most of which feature FPS games – he is very praising of the technique, stating at one point, *"Oh my god. Oh that's satisfying ... it feels amazing. Like the first piss in the morning."* One explanation for HAL's acceptance is the design decision to *augment* traditional controls, rather than replace them. As one player noted, *"I'd rather see more touches like this that augment controls in hardcore games than having control schemes that are all-or-nothing in the motion control department."* In another example, a gamer appealed to other players: *"…you do realise that this isn't so much motion controls so much as an addition to a traditional control scheme on the PC?… If you're a PC fan, you may actually want to check it out."*

Although the response to HAL was mostly favorable, several recurring concerns were noted. For example, several commentators feared that the use of head tracking could cause neck strain. This *perceived* issue was, however, scarcely reported by our study participants, with only three noting any *actual* strain, one of whom stated; *"My neck hurts, but in a good way!"* Although the study did not explore the use of head tracking over the very long durations that some players claim to play for, encouragingly, none of the 30 participants who played for more than an hour complained of any neck strain. Another health concern pertained to HAL's ability to induce feelings of motion sickness. As one participant stated, *"Dizziness! My main complaint is that the head tracking almost made me sick!"* This comment was reflected by a few other participants, with some specifically attributing these feelings to the handy-cam effect: *"Possibly being rid of the camera shake. It makes me dizzy fast."* Notably, these concerns were not raised in our previous focus group studies [23]. With only a small percentage of the participant sample raising the issue, its occurrence may be a result of the high probability of even a low prevalence event happening amongst a very large participant sample. The issue appears to have some precedent in the literature; with the work of [19] finding the inclusion of visual jitter can increase the realism of a scene at the cost of hastening the onset of motion sickness. The experiential benefits of the handy-cam were also noted to have a tradeoff with playing performance. One participant noted, *"I liked head tracking, it added a sense of more realism and immersion into the game. But sadly it was kind of frustrating to use since my accuracy greatly decreased."*

6 Head Tracking Guidelines

To help summarize many of the findings discussed in the previous section, we present a set of head tracking guidelines. These guidelines extend upon those presented in our earlier work [23] to reflect the insights we have gained in having examined head tracking within a real-world setting. Their intention is help other researchers with the design and creation their own head tracking techniques.

---

[2] `http://www.joystiq.com/2009/07/01/a-peek-and-head-tilt-into-the-future-head-tracking-in-fps-gam/`, *accessed Apr '13.*

Provide Motivation and Alternatives: A game experience designed around the use of head tracking and its benefits (i.e. the ability to perform partial movements alongside traditional controls) will see its use increase (see Sec. 5.2 for more details). Despite this, potentially quite a large proportion of players (up to a quarter) will still not use head tracking, either through choice (e.g. for playing performance reasons) or circumstance (e.g. inadequate lighting), so proper consideration still needs to be given to non-gestural alternatives.

Make It Adaptable: A home setting introduces a *very large* amount of variation with regards to the tracking quality and hardware performance. The following can help make a technique more consistent: convert the user's absolute head position into one relative to a mean position, use time-based (rather than sample-size-based) logic, make the logic more conservative under less-than-desirable conditions, and clamp the head data to a desirable range (see Sec. 3).

Not All Head Movements Are Equal: Avoid using the pitch or yaw axes for critical controls as these are noisy and, at extremes, make observing the screen awkward. By contrast, horizontal and roll movements are natural to perform and are more reliably tracked (see Sec. 5.2). Experimenting with the tracking software will help you determine which movements (i.e. axes and speeds) are better tracked than others [23].

Know Your Audience: Techniques that utilize a natural movement will be better received, remembering that some movements might not be natural to all users (e.g. cocking your head to peer down the barrel of a gun - see [23]). Consider what properties attract a player to a game and whether your technique enhances these (e.g. the handy-cam effect could probably enhance the immersion of a horror game, but interferes in a performance driven game). If players are particularly skilled in using a traditional control scheme, try to avoid replacing them entirely (sec 5.3). Finally, only by observing players use your technique will you know whether its intention has been properly conveyed (sec 5.2).

Put the Health of the Player First: Avoid long awkward movements and quick snapping actions as these may place the player under duress [23], as well as reduce the perceived appeal of the game (sec 5.3). Be mindful that a technique that alters the view perspective may cause some players to feel nauseous. Consider making the player have to proactively enable a technique to use it (sec 5.3).

Guide the Player: Guide the player through the set up of the head-tracking system. Written instructions and demonstrative images may be ignored by some players, so consider having the system dynamically highlight less-than-desirable issues as they arise (sec 5.1).

7 Conclusion

The new contribution of this research is two-fold. First, we extended and refined our earlier work [23]. This development required significant effort to modify the earlier prototype techniques to make HAL robust to the highly variable environment of home use. The home-use context set a rich background for assessing HAL, raising many challenges not found in the laboratory. These challenges led to new insights, which we have presented here. Foremost among these is the importance of the user's set-up in determining the effectiveness of the technique. Figure 4 is a simple illustration of the extent of the challenge facing widespread adoption of this technology – users, either intentionally or not, often present conditions that are almost insurmountable for reliable head tracking, and so coaching users to optimize their set-ups appears a necessary precursor to widespread use. The use of the home setting and the consequent large sample of participants also allowed us to identify a small but meaningful group of users who experience nausea while using our software, an result that did not occur during small-scale laboratory testing. The use of online recruitment and survey tools allowed us to amass a large amount of data quickly and easily, and our analyses revealed structure that showed that users were enthusiastic about head tracking and that even initially-skeptical users were swayed by its effectiveness in immersing them in the game. Head tracking did not produce an immediate performance boost – in fact, performance declined, perhaps because of users' unfamiliarity with the technique – although there was evidence of improvement with learning. This encourages us to consider a future longitudinal study to examine the long-term effectiveness of head tracking in home gaming.

Acknowledgements. The authors would like to thank Seeing Machines for the use of the FaceAPI software. No commercial funding was received for this research.

References

1. Andreasen, M., Nielsen, H., Schrøder, S., Stage, J.: What happened to remote usability testing?: an empirical study of three methods. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1405–1414. ACM (2007)
2. Bohemia Interactive Studio: ARMA 2 (2009), `http://www.arma2.com` (accessed April 2013)
3. Bradski, G.: Computer vision face tracking for use in a perceptual user interface (1998)
4. Breiman, L., Friedman, J., Olshen, R., Stone, C., Steinberg, D., Colla, P.: CART: Classification and regression trees. Wadsworth, Belmont (1983)
5. Brown, B., Reeves, S., Sherwood, S.: Into the wild: Challenges and opportunities for field trial methods. In: Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems, pp. 1657–1666. ACM (2011)
6. Bubka, A., Bonato, F.: Natural visual-field features enhance vection. Perception 39(5), 627–635 (2010)
7. Cooper, S., Khatib, F., Treuille, A., Barbero, J., Lee, J.: Predicting protein structures with a multiplayer online game. Nature 466(7307), 756–760 (2010)
8. Costanza, E., Giaccone, M., Küng, O., Shelley, S., Huang, J.: Ubicomp to the masses: a large-scale study of two tangible interfaces for download. In: Proceedings of the 12th ACM International Conference on Ubiquitous Computing, pp. 173–182. ACM (2010)
9. De Barros, P.G., Lindeman, R.W., Loughlin, T.J.: Characterizing Head Movement in First-Person Games. Poster, ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games (2009)

10. Evans, J.R., Mathur, A.: The value of online surveys. Internet Research 15(2), 195–219 (2005)
11. Gkikas, K., Nathanael, D., Marmaras, N.: The evolution of FPS game controllers: How use progressively shaped their present design. In: 11th Panhellenic Conference on Informatics, pp. 18–20 (2007)
12. Gorodnichy, D., Malik, S., Roth, G.: Nouse 'Use your nose as a mouse'-a new technology for hands-free games and interfaces. In: Proc. Intern. Conf. on Vision Interface, pp. 354–361 (2002)
13. Hartson, H., Castillo, J., Kelso, J., Neale, W.: Remote evaluation: the network as an extension of the usability laboratory. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 228–235. ACM (1996)
14. Kittur, A., Chi, E.H., Suh, B.: Crowdsourcing user studies with Mechanical Turk. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 453–456. ACM (2008)
15. Lombard, M., Ditton, T.: At the heart of it all: The concept of presence. Journal of Computer-Mediated Communication 3(2) (1997)
16. Microsoft Corporation: Kinect (2010), http://xbox.com/kinect (accessed April 2013)
17. Murphy-Chutorian, E., Trivedi, M.M.: Head Pose Estimation in Computer Vision: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence 31(4), 607–626 (2009)
18. NaturalPoint Inc.: TrackIR (2001), http://www.naturalpoint.com/trackir (accessed April 2013)
19. Oman, C.M.: Motion sickness: a synthesis and evaluation of the sensory conflict theory. Canadian Journal of Physiology and Pharmacology 68(2), 294–303 (1990)
20. Palmisano, S., Gillam, B.J., Blackburn, S.G.: Global-perspective jitter improves vection in central vision. Perception 29(1), 57–67 (2000)
21. Robinson, K.M.: Unsolicited Narratives from the Internet: A Rich Source of Qualitative Data. Qualitative Health Research 11(5), 706–714 (2001)
22. Seeing Machines: faceAPI (2001), http://www.faceapi.com (accessed January 2013)
23. Sko, T., Gardner, H.J.: Head tracking in first-person games: Interaction using a web-camera. In: Gross, T., Gulliksen, J., Kotzé, P., Oestreicher, L., Palanque, P., Prates, R.O., Winckler, M. (eds.) INTERACT 2009. LNCS, vol. 5726, pp. 342–355. Springer, Heidelberg (2009)
24. Sko, T., Gardner, H.J., Martin, M.: Non-parametric Decision Trees for Online HCI. To be Presented at: The 2013 ACM Annual Conference on Human Factors in Computing Systems. ACM (2013)
25. Stellmach, S., Dachselt, R.: Designing gaze-based user interfaces for steering in virtual environments. In: Proceedings of the Symposium on Eye Tracking Research and Applications, pp. 131–138. ACM (2012)
26. Sutherland, I.: A head-mounted three dimensional display. In: Proceedings of the Fall Joint Computer Conference, Part I, pp. 757–764. ACM (1968)
27. Wang, S., Xiong, X., Xu, Y., Wang, C., Zhang, W., Dai, X., Zhang, D.: Face-tracking as an augmented input in video games: enhancing presence, role-playing and control. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1097–1106 (2006)
28. Willow Garage: OpenCV (2000), http://opencv.willowgarage.com (accessed April 2013)
29. Wood, R., Griffiths, M., Eatough, V.: Online data collection from video game players: Methodological issues. CyberPsychology & Behavior 7(5), 511–518 (2004)

Tracking Eyes in Service Prototyping

Monchu Chen[1] and Veraneka Lim[1,2]

[1] Madeira Interactive Technologies Institute, Portugal
[2] Eindhoven University of Technology, Netherlands
monchu@m-iti.org, v.lim@tue.nl

Abstract. A mobile eye tracker was used to collect viewing behavior in a mixed reality immersive Cave Automatic Virtual Environment (CAVE) environment to evaluate a design concept of a tourist information office. The synthetic office consists of physical artifacts and virtual contents projected onto three walls of a room-sized cube. A Think Aloud study was conducted with both a goal-oriented condition and a free-browsing condition while subjects wearing the eye-tracker. Multiple Augmented Reality markers were used to reconstruct gaze positions in the coordinate system of the real environment. Gaze points were later aggregated to create heat maps, which were used as textures for a computer 3D model replication of the synthetic tourist office. The interactive visualization of the 3D heat map showcases different viewing patterns for different conditions. The insights suggest the combination of eye-tracking and mixed reality environment to be a valuable tool for prototyping service design of similar kinds.

Keywords: Eye-Tracking, Service Design, Experience Prototyping, Heat Map.

1 Introduction

Eye-tracking technique has been applied to various areas such as psychology, human factors, marketing research, etc. It has been particularly successful in the usability evaluation of human-computer interface design [1]. An eye-tracker allows researchers and designers to follow the eye gaze of a user, providing insights to how a user perceive and interact with an interactive system. Knowledge about when and where customers target their visual attention and how this is affected by certain aspects of the service or environment is expected to aid service designers in understanding customers' mental processes under circumstances like information processing and its influence on decision making. Several eye movement analysis metrics have been proved to be useful for desktop scenarios (e.g. a web site design), however, little have been applied to ubiquitous computing environments (e.g. a service design). Therefore, this paper proposes to depict customer experiences while using a service through the study of eye gaze.

In the last decades, most research on visual attention was primarily based on two-dimensional (2D) stimuli and static scenarios, regarding both observer and content (e.g. websites and interfaces). With the increased interest in interactions of people in

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 264–271, 2013.

natural settings and mobile uses of interaction technology, current advancements in mobile eye tracking systems now allows the observer to freely interact physically in a three-dimensional (3D) scenario, providing the means to assess eye gaze and visual attention and understand mental processes in arbitrary 3D environments (e.g. driving simulator and virtual reality). This approach answers questions on how to improve the guidance of visual attention afforded by 3D environments with multiple levels of depth. Research on visual attention, however, has still mainly targeted 3D objects and 3D virtual reality [2, 3, 4]. Limited to none research has focused on visual attention in Service Design, which is a new, holistic, multi-disciplinary and integrative field [5]. One reason why eye tracking hasn't been applied to Service Design might be related with the issue of eye gaze data collection and representation in 3D environments for an effective interpretation of eye tracking results.

A common technique for investigating visual attention is the aggregation and representation of gaze target positions in a scan path or a superimposed attention map, also commonly referred to as heat map [6,7,8]. Scan paths are drawn as linearly connected dot sequences, depicting raw gaze points (x, y, t) or a subset of those points that satisfy criteria for labeling as fixations, the relatively stationary component of the eye movement. Heat maps provide a much cleaner depiction of aggregate gaze by combining gaze fixations from multiple viewers. These two representations of gaze target positions are suitable for 2D stimuli, but as eye tracking studies typically comprise a time consuming frame by frame analysis of captured screen recordings with superimposed attention maps, the reduction of 3D scenarios to 2D is likely to lead to data loss without assessing the hypotheses on 3D stimuli too [2]. Although, several approaches has been done recently to transfer the notion of attention maps from 2D content to 3D content [3,6,7,9] they all have targeted only 3D objects and 3D virtual environments. Beside these, heat maps have received relatively little attention and have remained unchanged since their introduction. Hence, visual gaze analysis in real 3D environments lacks methods and techniques for aggregating attentional representations.

2 The Study

In this study, a 3D cave automatic virtual environment (CAVE) has been set up for prototyping the service design of a tourist information office space (see figure 1). A CAVE is an immersive virtual reality environment where projectors are directed to three, four, five or six of the walls of a room-sized cube, allowing the exploration of users' experiences in a more realistic but controlled space. Users' viewing behaviour will be collected with the Tobii [10] mobile eye tracker while experiencing the simulated service. Eye gaze patterns will then be presented on a 3D heat map for understanding customers' experiences while in a service. The purpose of this study is in twofold. First, as Experiential Prototyping lacks quantitative methods for gaining insights in users' experiences, the present study, introduces the use of eye gaze as a direct measurement into users' mental processes. Second, an implementation using Augmented Reality (AR) markers has been proposed for representing customers' visual attention with actual distances in a CAVE, while experiencing the service. The result, a 3D attention heat map, is expected to be a valuable tool for Service Prototyping.

2.1 Apparatus and Participants

A semi-virtual CAVE of a tourist information office was simulated with 3 large projection screens and real physical objects (an information kiosk, a table with brochures and a TV screen for showing advertisements) (see figure 1). The side screens were projected with a video (including audio) and static images with text variating in information content. The center screen was projected with a dynamic synthetic customer representative showing closed non-verbal behavior (e.g. on the phone or working on the computer).

The CAVE was used as an experience prototyping platform. 5 participants, composed of students and researchers from the University, were recruited for this study. The mobile eye-tracker, developed by Tobii [10], provided the technical basis for an analysis of visual attention during the mobile exploration in the CAVE (see figure 1).

Fig. 1. The simulated tourist information office

2.2 Augmented Reality (AR) Markers

Augmented Reality markers are small printable black and white shapes that interact with software downloaded to a computer equipped with a webcam or a smart phone with camera to create 3D images overlaid on the real world image seen on screen. A total of 16 AR markers were located strategically within the CAVE near visual targets. The set-up was designed so that at least 1 tag was always visible when participants looked through the Tobii Glasses.

2.3 Procedures

All participants were asked to enact a tourist information office visit by going through two tasks, an open task condition and a goal directed task condition. Two different tasks were chosen in order to create the possibility of getting two different types of behavioral approaches and gaze patterns. In the open task condition participants were given an absence of specific plans. In the second task condition they were given a presence of specific plans; they were asked to imagine the following “You heard a tourist talking about their great time during their hike from x to y. You are interested in this hike as well as other popular hikes and would like to know how to get there, if there is a possibility to camp and any other information about the trail. You also want to know about the possibilities to do whale watching. You have to select one for tomorrow’s activity”.

Each task, which lasted no longer than 5 minutes, was explained verbally to participants before entering the CAVE. After the task was explained they were asked to enter the CAVE and scan through the environment in order to see what they want to do, see or know more about. All participants were asked to think aloud as much as possible and say out loud as soon as they have decided what they want to approach. This was also an indication of the end of the task. For technical limitations regarding the mobile glasses and AR tags, participants were assigned to first walk to a starting point in the CAVE, indicated with a round symbol on the floor. They were also asked to stay within a particular area when in the CAVE, which was also indicated on the floor.

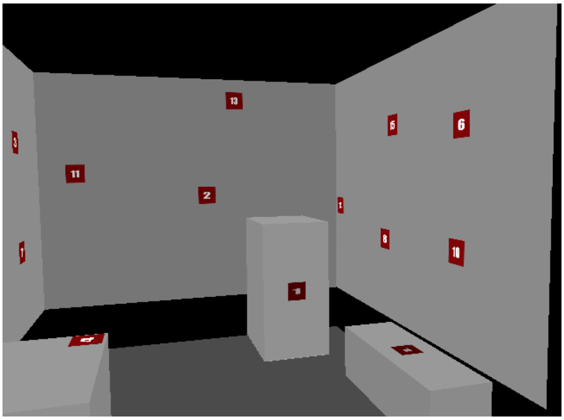


Fig. 2. A visualization of the AR markers in the virtual space

2.4 Implementation

After data collection, the first step was to re-create the tourist information office in virtual space using Quartz Composer Development Environment available in Mac OSX, where interactive 3D navigation and real-time visualization is possible. Figure 2 shows the 3D virtual space of the tourist information office tagged with AR markers, which were used as anchor points for projecting gaze points. Gaze positions were converted from 2D into 3D space for each AR marker individually (see figure 3). A video layer was first captured by the scene camera on Tobii Glasses and a gaze layer with eye gaze position was acquired by Tobii Studio Gaze positions in scene camera coordinate system. Since Tobii Glasses is an off-line eye tracker, post-processing is required after experiments. Gaze data and video streams from scene camera were analyzed and exported from Tobii Studio before sending to Quartz Composer for further analysis. Then, the AR markers are detected and recognized in the Quartz Composer. The marker position and rotation in scene camera coordinate system were extracted. The origin was then shifted to the center of the AR marker and the video stream from the scene camera as well as the gaze points are translated to the AR marker's coordinate system. A 3D rotation transformation was performed to adjust the perspective so that the norm vector of the maker is perpendicular to the display. Finally, the eye gaze position was transformed to the appropriate position related to the marker in the virtual tourist office.

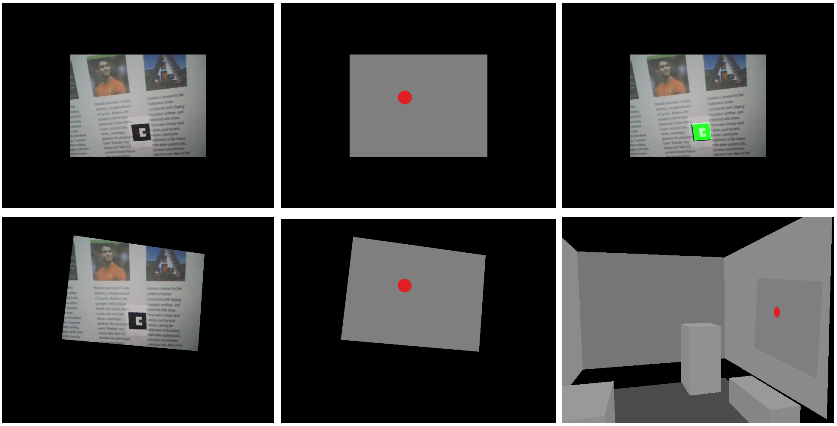


Fig. 3. A video stream captured by the scene camera on Tobii Glasses (top left); A gaze layer with eye gaze position acquired by Tobii Studio Gaze positions in scene camera coordinate system (top middle); AR markers (marker id# 10) detection and recognition of marker position and rotation in scene camera coordinate system (top right); 3D rotation transformation (bottom left); Eye gaze position transformation to the appropriate position (bottom middle); gaze position converted from the marker coordinate system into the coordinate system of the tourist office (bottom right).

Since the positions of all 16 AR markers in the tourist information office coordinate system were pre-measured in the coordinate system of the tourist office, references were known for calculating gaze points in 3D space. Gaze points are further converted from the gaze position in the marker coordinate system into the coordinate system of the tourist information office. Gaze points were aggregated over the course of the experiment, and across all subjects. The intensity of the luminance represents the frequency of the visual attention. Hence, the "hot" spots in the virtual tourist office indicate areas that attract customers more attention.

3 Results

Participants' verbal comments during the think aloud, on the spot observations of participants behaviour and video material were analyzed qualitatively. Overall, in the open task condition, participants seem to divide their attention to different targets within the CAVE in a more balanced manner spending the same amount of time on fixating towards a target as well as moving to other targets. They made more movements and put more effort in trying to cover all information available in the tourist information office.

In the goal directed task condition, participants seem to scan quickly through all targets until they find what they were searching for. They spend more time on fixating toward a target of interest and less time for moving to other targets.

Figure 4 shows the 3D attention heat map of the aggregated eye gaze for all participants separately for both conditions. It seems that in the goal directed task condition, the heat map is more focused and concentrated supporting the qualitative findings. The 3D heat map, however, provides specific details on what has been targeted. The intensity of the heat map reflects the quantity of visual attention of a given position. The by-product of the process provides the access to the raw data of gaze positions in 3D space. With further defining Area of Interests, other quantitative analysis such as fixation count, fixation duration, percentage fixed, etc. could be easily obtained.

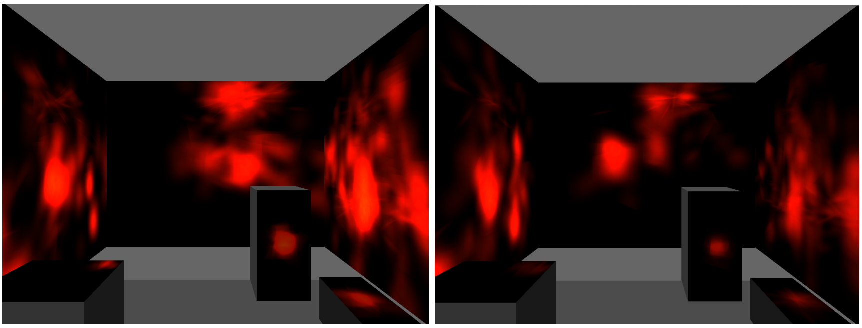


Fig. 4. A visualization of the 3D heat map in the virtual space for the open task condition (left) and goal directed task condition (right)

4 Discussion and Conclusion

The interest in exploring and developing experiences in HCI is growing, while often designers of interactive systems go beyond their limits when using tools to explore and communicate what it will be like to interact with the things they are designing. A popular tool to understand, explore or communicate what it might be like to engage with the product, system or space that is being designed is Experience Prototyping. This tool has been of importance in the application of Service Design. Hence, Experience Prototyping is qualitative in nature and characterized with an indirect approach into customers' experience or view towards products or services. Therefore, one of the purposes of this study is to introduce a quantitative method for gaining insights in users' experiences, by looking at eye gaze as a direct measurement into users' mental processes.

Although current advancements in mobile eye tracking systems allows the observer to freely interact physically in a 3D scenarios, providing the means to assess eye gaze and visual attention and understand mental processes in arbitrary 3D environments, eye tracking hasn't been applied to Service Design yet. Visual gaze analysis in real 3D environments also seem to lacks methods and techniques for aggregating attentional representations, which directs us to a second purpose of this study; an implementation of representing customers' visual attention with actual distances, while experiencing the service using AR Markers.

A mixed reality CAVE was set up for prototyping the service design of a tourist information office space and users' viewing behaviour was collected with a mobile eye tracker while in a service. Eye movements were recorded and visualized virtually on a 3D attention heat map. The different viewing patterns for different conditions, concludes the 3D heat map to be a valuable tool for prototyping service design. The 3D heat map provides specific details on what has been targeted compared with qualitative methods. The proposed paradigm is independent to the brand, type and model of mobile eye-trackers. Additional efforts can be made to port the software to support other mobile eye-trackers in order to benefit service design community.

Some limitations were found during this study. First, the slow shuttle speed of the scene camera in the eye tracker resulted in blurry images when head movements occurred, which makes it more difficult for AR marker detection. Second, there were problems with the marker distribution within the CAVE. When more markers are used, indicating higher density, the performance improves but obscures the scene. The use of IR markers could be a solution for future studies. In additional to invisible markers, object and scene recognition should be considered as well when tracking in a more realistic scenario where planting markers to cover whole physical space is not feasible. Also, the implementation requires manual operation. Therefore, an eye gaze proxy could be useful once online access to gaze data on the mobile tracker is possible. Future work should focus on more empirical studies to validate the usefulness of the 3D heat maps as well as other new ways of quantitative analysis of visual attention that can be applied to Service Design.

Acknowledgments. The study was done in Logica Service Design Lab with the supports from Knowledge+ incentive system in Madeira, Portugal. We would also like to thank all the volunteers who participated in the study during data collection.

References

1. Jacob, R.J.K., Karn, K.S.: Eye tracking in Human-Computer Interaction and Usability Research: Ready to deliver the Promises. In: Hyönä, J., Radach, R., Deubel, H. (eds.) The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research, pp. 573–605. Elsevier Science, Amsterdam (2003)

2. Pfeiffer, T.: Measuring and Visualizing Attention in Space with 3D Attention Volumes. In: Proceedings of the Symposium on Eye Tracking Research and Applications, March 28-30, pp. 29–36. ACM, New York (2012)
3. Stellmach, Nacke, L., Dachselt, R.: 3D attentional maps: Aggregated gaze visualizations in three-dimensional virtual environments. In: Proceedings of the International Conference on Advanced Visual Interfaces, pp. 345–348. ACM (2010)
4. Stellmach, S., Nacke, L., Dachselt, R., Lindley, C.: Trends and Techniques in Visual Gaze Analysis. In: Proceedings of the 5th Conference on Communication by Gaze Interaction - COGAIN 2009: Gaze Interaction for Those Who Want It Most, pp. 89–93. The COGAIN Association, Lyngby (2009)
5. Moritz, S.: Service Design, practical access to an evolving field. Koln International School of Design, London (2005)
6. Pfeiffer, T.: Tracking and Visualizing Visual Attention in Real 3D Space. In: Haack, J., Wiese, H., Abraham, A., Chiarcos, C. (eds.) Proceedings of the KogWis 2010, pp. 220–221. Universittsverlag Potsdam, Potsdam (2010)
7. Spakov, O., Miniotas, D.: Visualization of Eye Gaze Data using Heat Maps. Electronica and Electrical Engineering 2(74), 55–58 (2007)
8. Velichkovsky, B.M., Hansen, J.P.: New technological windows into mind: There is more in eyes and brains for human-computer interaction. In: CHI 1996: Human Factors in Computing Systems, pp. 496–503. ACM Press, NY (1996)
9. Pfeiffer, T.: Understanding Multimodal Deixis with Gaze and Gesture in Conversational Interfaces. Berichte aus der Informatik. Shaker Verlag, Aachen (2011)
10. Tobii Glasses Eye tracker, `http://www.tobiiglasses.com`

Did We Miss Something? Correspondence Analysis of Usability Data

Stanislaw Zabramski[1] and Wolfgang Stuerzlinger[2]

[1] Department of Informatics and Media, Uppsala University, 751 05 Uppsala, Sweden
stanislaw.zabramski@im.uu.se
[2] Department of Computer Science and Engineering, York University, Toronto, Ontario, M3J 1P3, Canada
www.cse.yorku.ca/~wolfgang

Abstract. We have applied a multivariate exploratory technique called Correspondence Analysis (CA) to create and analyze a model of the dataset of experiment results. The dataset originates from a comparative usability study of tracing with the use of mouse, pen, and touch input and contains both categorical and continuous data – i.e. results of questionnaires and task measurements. CA allowed to visually and numerically assess the main variables in the dataset and how they interact with each other. In our study, pen input had the best measured performance and was preferred by the users. Touch input was the least accurate of all input methods tested but it was preferred by users over mouse especially in the conditions lacking of visual feedback of drawing. CA helped to detect that secondary effect even though it cannot be explained by the performance results alone. The importance of the influence of user's previous experience is also noted. We conclude that CA helped to identify all major phenomena known from previous studies but also was sensitive to minor and secondary effects, what makes it a well suited method to quickly evaluate usability data.

Keywords: shape, freehand, tracing, drawing, mouse, pen, stylus, touch, evaluation, comparison, error, measurement, subjective.

1 Introduction

Usability of any technical solution can be perceived from the pragmatic point of view that focuses only on "getting things done". However, there is a variety of users' beliefs, preconceptions, preferences, perceptions, and physical and psychological responses that affect the whole experience before, during and after use of the technique tested. Therefore the subjective and objective approaches should be used in usability evaluations and the results should be integrated into one consistent picture [1]. Even though subjective perceptions of usability are generally not correlated with objective measurements [2] their comparison is done rarely in practice [3]. The reason for that might be that it is hard and time consuming to get a holistic view on the experiment

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 272–279, 2013.

from results obtained from a number of different methods of data analysis suitable for particular types of data.

Correspondence Analysis (CA) is a multivariate exploratory technique similar to Principal Component Analysis (PCA). In this study we apply CA and analyze objective measurements of users' performance i.e. task time and error, together with subjective evaluations of users' satisfaction in shape-tracing task.

Artistic drawing involving computer as a tool may be negatively influenced both by inaccuracies in the input method and by spatio-temporal constraints imposed on the user e.g. by task formulation. Unfortunately, there is no suitable model to describe spatially and temporally unconstrained freehand drawing when the path is unknown a priori and typical user errors are not characterized mathematically. Therefore, an exploratory study based on a standardized method [4] was performed to comparatively evaluate three popular input methods (mouse, pen, and touch input) in a simple shape-tracing task [5]. Previous research points to a positive influence of visual feedback. Therefore, the visual feedback of drawing was controlled in analyzed study through the presence or absence of solid black ink trace.

We use CA in an attempt to get a quick and simplified view on the experimental data-set and use it to evaluate if drawing performance can be a predictive criterion of users preferences [6] towards input methods tested with controlled visual feedback.

2 Procedures and Methods

We used CA to analyze the data-set of results from experimental study that aimed to evaluate the influence of mouse, touch and pen input on user performance in an *unconstrained* free-hand tracing task [5]. It also included the user experiences and preferences expressed towards these devices.

2.1 Experiment Design

The experiment featured a mixed design. Sixteen student volunteers (4 females) with average age 24 (SD = 2.3) were selected through convenience sampling from the local university campus. All except one were right-handed and had normal or corrected to normal vision. Half of the participants were randomly assigned to the visible drawing line condition and the other half had no feedback of drawing. Each participant used the mouse, pen and touch input on a Windows7 based HP Touchsmart TM2-1090eo Tablet PC to draw over a semi-randomly generated contour shape displayed on the screen once with every input device. Devices were assigned in a counterbalanced order in a within subject design. They were required to use the same starting point and direction to draw over the same nonsense closed contour shape presented, using a single continuous and uninterrupted stroke. Timing data was collected for each task. User error was calculated via a method that computes the average deviation of the user-generated shape from the original version [4, 5].

2.2 Questionnaires

All the participants filled pre-test questionnaires and provided basic screening information about demographics as well as their familiarity with the investigated computer input methods in the last 3 years.

After the experiment, post-test questionnaires were used to collect information about participant experience and opinions regarding three following aspects:

- input device of preference (statement S1)
- perceived ease of use (statements S2 and S4)
- perceived learnability (statements S3 and S5)

To avoid potential distortions originating from acquiescence bias the scale was balanced by offering equal number of positive and negative statements regarding the most important aspects [7]. The users' ratings of negative statements were later inverted and the final ratings then computed as the median of the responses to the paired questions (S2&4, S3&5). The questionnaire included:

- S1 - I think that I would like to use this input device frequently for sketching.
- S2 - I thought that the input device was easy to use.
- S3 - I would imagine that most people would learn to sketch with this input device very quickly.
- S4 - I found the input device very cumbersome to use.
- S5 - I imagine I would need to practice a lot before I could sketch efficiently with this input device.

Participants were asked to evaluate five statements on a symmetric Likert-like psychometric scale. The choices offered were: strongly disagree, somewhat disagree, indifferent, somewhat agree, strongly agree. Every statement was evaluated for each input method separately.

2.3 Correspondence Analysis

CA identifies a set of synthetic variables (factors) that summarize the original data set. In this study, CA is expected to reveal frequency-based associations between the categorical data from the questionnaires and the data from the measurements [8]. CA creates a model for data with fewer factors by decomposition of inertia describing the degree of variation or eigenvalue. Similarly to PCA, the calculation gives the factors and correlations (loadings) with each factor for each variable. The optimal number of factors can be decided with the scree-test or the Kaiser criterion. Then significance tests can be used to indicate that the occurrence of certain values is significantly more common than chance.

CA also enables the visualization of associations between variables in question, to simplify their identification, and to isolate relations between them. That can be done through two-dimensional configuration bi-plots, which show the relative proximity of data points and depict degrees of correlation and variation (inertia). Each dimension of a plot explains a certain percentage of the variation in the data. Normally, the first two dimensions capture a large percentage of the variation [8].

3 Results

3.1 Pre-test Questionnaires

The participants reported their average daily experience: 8.2h (SD = 5) of with computers in general, 3.7h (SD = 4.7) with a computer mouse, 0.8h (SD = 1) with pen devices (incl. mobile phones), and 3.3h (SD = 5.3) with touch devices (incl. mobile phones). These values have been called *experience* and were used in the analyses.

3.2 Time and Error

The ANOVA of task time and user error data showed no main effect for the visibility of visual feedback during drawing, nor any interaction between feedback visibility and input device used [5]. There was no significant difference between the visibility of feedback conditions either for time (Cohen's d = 0.3) or error (Cohen's d = 0.14). Power analysis revealed that in order for an effect of this size to be detected (90% chance) as significant at the 5% level, a sample of 238 and 1111 participants would be required for time and error respectively.

However, there was a main effect of input device on task time and a post-hoc analysis identified significant differences between touch input and mouse. The mean times were: mouse = 40 sec. (SD = 16.6 sec.), pen = 30.1 sec. (SD = 15.3 sec.), and touch = 21.5 sec. (SD = 9.3 sec.). User error data indicated also a main effect of input device. A post-hoc analysis showed that mouse and pen are not significantly different. The mean errors were: mouse = 5.84 px (SD = 2.21 px), pen-input = 4.98 px (SD = 1.92 px), touch-input = 7.95 px (SD = 2.94 px), with the grand mean error = 6.25 px.

3.3 Post-test Questionnaires

All the questionnaire responses were coded between 1 and 5 reflecting the level of expressed agreement or disagreement on an ordinal scale. Some of the scales were inverted to deal with acquiescence bias. We illustrate this inversion by adding a "-" sign in front of the statement number (e.g. -S4). The paired statements, S2&-S4 and S3&-S5, were compared to check for acquiescence bias.

Polychoric correlation coefficients (pcc) were calculated and used in the analyses. To investigate the significance of potential differences between groups of responses a Wilcoxon matched-pairs test was used.

We found that the overall replies to statements S2 and -S4 for all input methods and all feedback visibility conditions have strong positive correlation (pcc=0.68) with insignificant differences between their medians. Similarly, in case of statements S3 and -S5, strong positive correlation was observed (pcc=0.74), but this time with significant differences between their medians ($p<0.05$).

Figure 1 presents the summary of responses to post-test questionnaires.

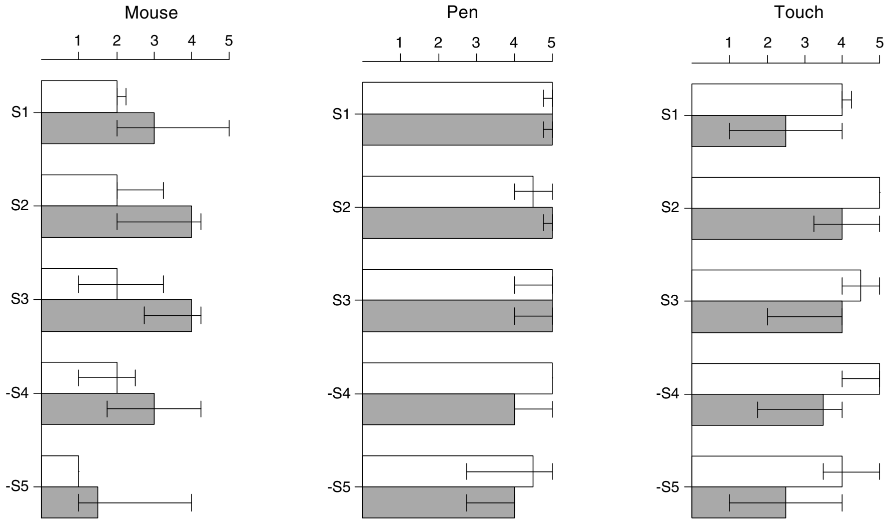


Fig. 1. Median responses to post-test statements for each input method and visibility of feedback mode *(white bars for feedback "off", grey bars for feedback "on")*. Whiskers represent the interquartile range.

In following analyses the median value of replies to S2 and -S4 is referred to as *ease of use,* and the median value of replies to S3 and -S5 as *learnability*. Replies to S1 are termed *preference*. The correlation matrix of responses of all these three aspects illustrated positive correlations to each other (pcc>0.5).

3.4 Correspondence Analysis

For extracting common explanatory factors a CA was calculated and analyzed. The first analysis compared the experience in using various input methods against the measurements and the post-test results, such as preferred input device and perceived ease of use.

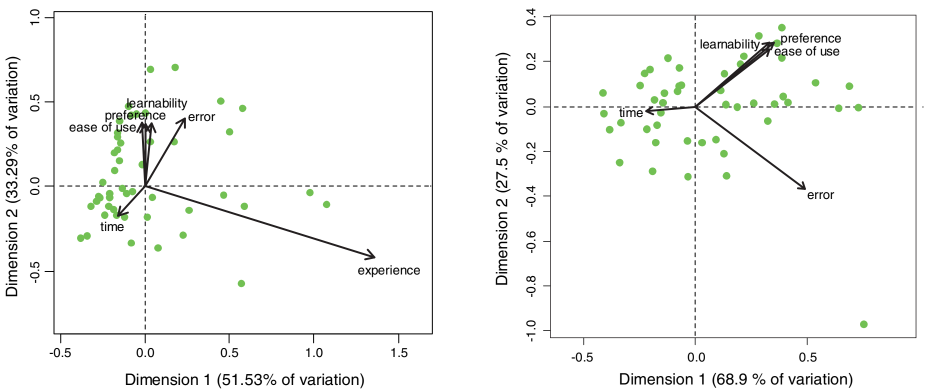


Fig. 2. Bi-plots from Correlation Analyses with *(left)* and without users' *experience* as variable in the dataset *(right)*

The bi-plot generated from the CA (see Figure 2 - left) shows that *experience* has a strong influence on all other variables. It shows potential strong relation to task time, but neither the correlation nor regression analysis gave any significant or consistent results. The *experience* factor was removed from data for the following analyses to better contrast the remaining variables (see Figure 2 - right).

Some differences between the visibility of the drawing feedback conditions have been noticed in the post-test responses within each input device (see Figure 1). The bi-plot generated from the CA confirmed that observation (see Figure 3).

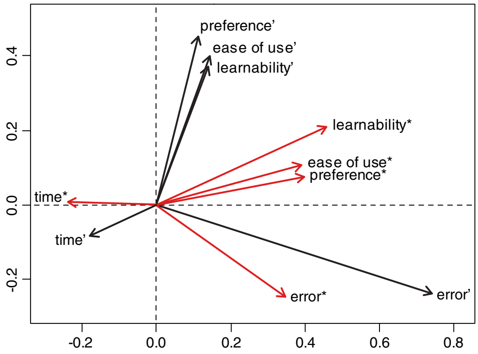


Fig. 3. Bi-plot from Correlation Analysis. The datasets with *(marked with apostrophes)* and without visible drawing feedback *(marked with asterisks)* were considered separately.

A Mann-Whitney-Wilcoxon test identified significant differences for *preference* and *ease of use* only for touch input ($p<0.05$). In case of *learnability* that effect was not so strong for touch but a significant difference has been found in case of mouse ($p<0.05$). Table 1 summarizes these results.

Table 1. Input methods ordered by users preference in each visual feedback condition for each aspect. "≥" marks insignificant differences and ">" marks significant differences ($p<0.05$) after Mann-Whitney-Wilcoxon test.

Aspect	Visual feedback	Input methods
preference	on	Pen > Mouse ≥ Touch
	off	Pen ≥ Touch > Mouse
	on and off	Pen > Touch ≥ Mouse
ease of use	on	Pen > Touch ≥ Mouse
	off	Touch ≥ Pen > Mouse
	on and off	Pen ≥ Touch ≥ Mouse
learnability	on	Pen > Touch ≥ Mouse
	off	Pen ≥ Touch > Mouse
	on and off	Pen > Touch ≥ Mouse

4 Discussion

The multidimensional dataset generated by an experiment is challenging to analyze. Thus, it is essential to use a suitable analysis method, such as CA, which can identify the most influential variables.

Previous experience strongly varied between participants due to the lack of homogeneity in the participant's group. As the most influential variable, it biased both the measurements and subjective factors, but unfortunately not in a way that can be easily interpreted in a meaningful way. E.g. experience did not correlate with error or task time. Therefore it was excluded from the analyses.

CA suggested to contrast time with error (see Figure 2 - right). This is the evidence that the well-known speed-accuracy trade-off [5] took place in this study.

Subjective evaluations did not show meaningful differences between the aspects of user satisfaction addressed in questionnaires and strongly correlated to each other. The pen supports better performance and overall results for *device of preference* identify the pen as the best choice for frequent sketching. It is followed by touch and mouse with a non-significant difference between the two. However, this difference changed to significant when no feedback of drawing was provided (see Table 1) making touch more preferred when the feedback was absent. That trend is also present in the case of perceived *learnability* and even more visible for perceived *ease of use*. None of these however, correspond to significant differences of performance [5].

5 Conclusion

Correspondence Analysis is a frequently used technique e.g. in linguistics, sociology, or biology. When applied in HCI, CA helped to quickly analyze the results and identified all major phenomena in the studied experiment that mostly have confirmation in previous research. But, CA was also sensitive to secondary effects that might be harder to notice when the results are spread among multiple methods of data analysis.

In the experiment analyzed here the lack of visual feedback significantly improved the overall users' attitude towards touch as an input method for drawing (see Table 1 and Figure 1) – while having no significant effect on performance. The result is opposite for *learnability* for the mouse. The pen stays an obvious choice for drawing task both from performance and preference point of view. However, we think that more attention should be directed towards user biases favoring touch input over mouse observed in cases lacking of visual feedback. Also, the huge influence of user's familiarity with these input methods is worth of being investigated even more. Potentially, technical properties of the input methods, such as directness, occlusion effect, or latency may also contribute to the observed mismatch between the performance results and subjective evaluations.

Obviously, we have to assume that the results of the experimental study used here are true only for the particular shape investigated. Nevertheless, CA appears to be an available [9] and convenient method that might be appropriate also for exploratory analyses of multi-dimensional data-sets with results from other kinds of studies involving mixed data-types, also with much higher sample size - what makes it worth being promoted more widely in HCI community.

Acknowledgments. I thank Dimitrios Gkouskos and Sarathkumar Neelakannan for their respective input.

References

1. ISO: 9241-210 Ergonomics of human-system interaction - Part 210: Human-centred design for interactive systems. ISO 9241 (2010)
2. Frøkjær, E., Hertzum, M., Hornbæk, K.: Measuring Usability: Are Effectiveness, Efficiency, and Satisfaction Really Correlated? In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2000, pp. 345–352. ACM Press, New York (2000)
3. Barkhuus, L., Rode, J.A.: From Mice to Men – 24 years of Evaluation in CHI. Alt.CHI 2007 (2007)
4. ISO: 9241-9 Ergonomic requirements for office work with visual display terminals (VDTs) - Part 9: Requirements for non-keyboard input devices. ISO 9241 (1999)
5. Zabramski, S.: Careless touch: A comparative evaluation of mouse, pen- and touch-input in shape tracing task. In: OZCHI 2011, pp. 329–332. ACM, Canberra (2011)
6. Nielsen, J., Levy, J.: Measuring usability: preference vs. performance. Communications of the ACM 37, 66–75 (1994)
7. Schriesheim, C.A., Hill, K.D.: Controlling Acquiescence Response Bias by Item Reversals: The Effect on Questionnaire Validity. Educational and Psychological Measurement 41, 1101–1114 (1981)
8. Greenacre, M.J.: Correspondence Analysis in Practice. Chapman & Hal/CRC (2007)
9. Nenadic, O., Greenacre, M.: Correspondence Analysis in R, with Two- and Three-dimensional Graphics: The ca Package. Journal of Statistical Software 20, 1–13 (2007)

Social Overlays: Collectively Making Websites More Usable

Tao Dong, Mark S. Ackerman, Mark W. Newman, and Gaurav Paruthi

University of Michigan, Ann Arbor, MI, USA
{dongtao,ackerm,mwnewman,gparuthi}@umich.edu

Abstract. Many small organizations lack the expertise and resources to conduct usability evaluations of their websites. Social Overlays, presented here, is a new system that allows a community of users to collectively improve their website.

Social Overlays enables end–users to identify and repair common user interface problems through creating "overlays" on web pages as part of their regular use, thereby improving usability while reducing the need for professional services. In short, Social Overlays harnesses the diversity of experience and ideas within a community to "crowd source" usability.

To evaluate Social Overlays, we examined whether a group of community members without any usability training could use Social Overlays to identify and repair UI problems on their medium–sized community's website. We found that they could. Community users were able to uncover a large number of UI problems and formulate reasonable solutions to the problems they identified. In addition, we compared Social Overlays to two standard ways of assessing website usability: expert inspection and usability testing. We found that Social Overlays users identified more problems, and their reported problems differed in useful ways from those found by the experts and the usability testing team.

Keywords: Usability, community, peer production, social computing.

1 Introduction

Many small organizations' websites need usability improvement. For organizations like charities, museums, and schools, their constraints on technical resources and usability expertise keep them from sufficiently making their websites easy to use. For example:

- TriCounty GoodDeeds is rolling out its website. But, its web designer and webmaster are volunteers, and there are no spare resources for usability tests. How can they create a web site that is usable and helpful for the charity's constituents?
- The Tree City municipal government's financial services group created a new website for travel reimbursement. It is unusable, since it is full of jargon and appears to be meant for accounting professionals. How can staff in departments help the financial group with their usability?
- The History Department at the University of the Midwest recently created a new website using a content management system. The department has a well-meaning

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 280–297, 2013.

web developer, who is an ex–computer science major. The students have to use the website to get information for courses, requirements, and school events, though they often feel the developer doesn't know how they actually use the website. How can the department create a usable, useful website?

At a high level, there are two standard approaches to evaluating usability on websites: usability testing and expert evaluation [13]. While these approaches have many variants, they all share the characteristic that usability experts play a critical role in determining the existence and nature of usability problems. An alternative approach is to solicit problem reports from users. Post–deployment usability approaches focus on collecting feedback from users in the field at the time the problem is encountered [7, 12]. However Chilana et al. found that few usability practitioners analyze or respond to such feedback [6]. Participatory techniques have been proposed for involving users in usability evaluation during the formative stages of system development as well [2, 11], though these techniques still depend on the participation of usability experts and product developers. In this paper, we investigate a novel approach to creating usable interactive systems: enabling community members themselves to collectively improve the system as part of their everyday interactions with it. Our approach is aimed primarily at websites built to serve small–to–medium sized organizations or communities—precisely the sites that often do not have the resources to hire usability professionals or implement frequent changes to a site. Our question, then, is how can we leverage the user community of a website to uncover and service usability problems?

We have designed and built a system called Social Overlays (SO). By harnessing the "wisdom of the crowd," SO allows a community to collectively evaluate and improve their website, without the need for formal usability methods or professional usability expertise. Using SO, community members can create "overlays," which effectively rewrite particular page elements (e.g., text, links, and tooltips), thereby improving the site's usability for subsequent visitors. SO also provides lightweight mechanisms for different community members to nominate potential problems, propose alternate fixes for the identified problems, and vote for the best solution. In short, SO harnesses the diversity of experience and ideas within a community to "crowd source" usability.

While, at a technical level, SO could be deployed on a wide range of sites, our initial focus is on a type of site that is particularly in need of and well-suited to the SO approach. Those are the sites that serve communities or organizations with a few dozen to several hundred members. While such communities are the ones most in need of a low-cost approach to usability improvement, we also expect them to be relatively cohesive and possessing members who have sufficient common ground in vocabulary, practices, and expectations [10]. In these communities, members often know one another, creating the grounds for altruism and self–policing [17]. In a word, those communities possess the desirable social properties that would allow the SO approach to be adopted most effectively.

To examine the viability of the SO approach, we conducted a study with thirteen members of a medium–sized academic community. We found that they were able to find and repair a large number of usability problems on the community's website. Moreover, compared with usability experts whom we asked to evaluate the same site and an external usability team who conducted a conventional usability test, the community members reported more problems. The problems they found differed in systematic but useful ways from those found using standard usability methods. Thus, our study results argue for the feasibility of the SO approach, at least for small-to-medium sized communities and organizations.

The contribution of this paper, then, is twofold:

1. We present a novel approach to collectively improving website usability by enabling the site's community to not only identify but also repair usability problems.
2. We provide evidence that our approach leads to usability improvements for important types of communities that are comparable to standard but more costly methods such as expert inspection and usability testing.

In addition, we discuss how this approach can be extended for larger or less socially cohesive communities and how this approach is situated in the literature.

2 System Description

The goal of Social Overlays (SO) is to allow users of a website to identify and repair usability problems during the course of their regular use of the site. We identified, through pre-studies, an initial set of requirements that SO must meet in order to accomplish this goal.

1. It should be easy for non-technical users to report and repair problems.
2. When a user is aware of a problem but does not know how to repair it, he/she should be able to request help from other community members or the webmaster.
3. Community members must be able to see the repairs made by other users and decide whether or not they prefer a peer-modified version to the original design of a UI element.
4. Community members and the webmaster need to be able to review and address requests made by other users.

In the rest of this section, we first walk through a scenario that illustrates the use of SO, and then describe the main features of SO.

2.1 Scenario

Chelsea, a master's student at the aforementioned history department, is looking up a class she is considering for next semester. On the page that allows her to search for courses, she finds herself annoyed by the weirdly labeled "Apply" button next to the course search field. She wonders, "Why couldn't they just put the word 'Search' on that search button?" Coming from a literature background, Chelsea often jokes that she is a "language snob."

Fortunately, Chelsea has a new browser extension called Social Overlays that allows her to revise the button's label. She does so, and sees the change immediately.

In more detail, Chelsea chooses the Text tool in the SO panel (see Fig. 1a). Now when her mouse hovers over a page element, it is highlighted with an orange dashed outline (see Fig. 1b). She clicks the "Apply" button that annoyed her on the page to invoke SO's element editing dialog box, where she enters the text string "Search" to change the button's label (see Fig. 1c). Clicking the "Save" button instantly applies the change for her.

Within an hour, Chelsea finds that her overlay (the alternate label for the button) has received thumbs-up from 5 other SO users in her department. After several days, her department's IT person notified her that the website has implemented her change permanently, because of strong user preference shown on SO.

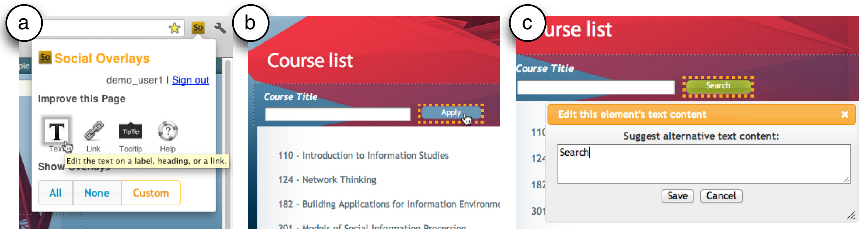


Fig. 1. Social Overlays has a three-step process of making modifications to an existing web page: a) select a modification type, b) select a page element, and c) specify the modification. For example, the user in this scenario renames a mislabeled button from "Apply" to "Search."

2.2 Overlays

In the above scenario, Chelsea repairs an unconventionally labeled button by creating an *overlay* using the Text tool provided by SO. Overlays are user–generated in–place modifications to existing web page elements. There are three types of overlays supported in the current version of SO: Text, Links, and Tooltips. Each type of overlay can be created and edited using tools provided by the SO extension panel (see Fig. 2).

As the scenario above has shown, the *Text* tool allows the user to revise inaccurate or unintuitive terms or languages used on buttons, links, or headers.

The *Link* tool allows the user to add a link to any element of a web page by entering a target URL. This can be useful for creating navigation shortcuts or pointing to additional information that might be helpful for a given task. In addition, the Link tool allows the user to edit the URL of a broken or outdated hyperlink.

The *Tooltip* tool allows users to create or rewrite tooltips (i.e., short messages that appear when an element is hovered over) that are attached to any page element. This tool allows a user to explain or clarify what an element does (or does not do) and how to make use of a feature on the site to his/her fellow users.

The Link tool and the Tooltip tool follow the same workflow as the Text tool to make overlays. A user can combine different types of overlays if needed. For example, the user can modify the text of a hyperlink, and also install a tooltip for it.

2.3 Help Requests

Although the changes that SO currently supports are relatively simple, some users might not feel confident enough to make a change. Additionally, some users might want a more sophisticated change that SO does not yet support. In such situations, a user can request help from the community or the webmaster using the Help button (see Fig. 2). The Help button allows the user to attach a message to the element that needs to be fixed or improved through a process similar to editing a button's label. If the request is not related to a particular element, the user has the option to submit a general request for the page by selecting the "bucket" at the bottom–right corner.

2.4 Indicators

When a page loads, SO shows page modifications made by the user community on the current page. A modified element will flash for half a second to help a user distinguish it from the original element, unless a particular overlay has been previously approved by the current user or by a certain number of other community members. Inspired by Edit Wear [8], SO uses the side margin of the web browser to provide lightweight visual indicators of prior community activity on the current page (see Fig. 2). Each indicator corresponds to an element that has community–generated overlays or requests, and they are visually aligned on the same line. When the user hovers over an indicator, its associated element will be highlighted. The color of the indicator shows the status of the element. A green indicator signals that there are existing overlays for the element, while a red indicator signals that there are active requests to fix this element. If an element has both overlays and unresolved requests, a yellow indicator is displayed.

If the user wants to see the original version of the element, clicking on the indicator will toggle between the original version and the community–enhanced version. He can also toggle between the original page and the community–enhanced page using the "Show Overlays" switch on the SO panel.

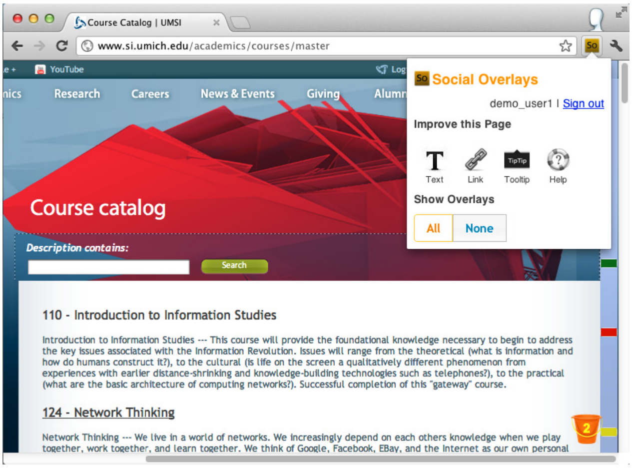


Fig. 2. The Social Overlays browser extension can be accessed by clicking its icon on the browser's toolbar. The icons from left to right represent the Text tool, the Link tool, Tooltip tool, and the Help button. The green, red, and yellow indicators on the right margin are used to indicate overlays, help requests, and both overlays and requests, respectively, made on a page element on the same horizontal line. The bucket at the right–bottom corner is a repository for page–wide issues.

2.5 Selection Rules

When multiple users have defined overlays for an element, one of them must be chosen to display. We implemented a voting mechanism that allows each user to indicate whether an overlay works for him or not. The most approved overlay for an element is automatically displayed unless the current user has approved a different one. In addition, if an overlay's number of votes reaches a community–specific threshold, it will no longer flash to draw the user's attention when the page loads, since it is likely to be working for her as well.

2.6 Implementation Details

We implemented SO as an extension for the Chrome web browser[1] that is coupled to a server–side application. Every time the user visits a new web page, the SO extension injects a set of JavaScript files into the current page to access and manipulate the page's HTML Document Object Model (DOM). With the full control of the target page's DOM, the SO extension enables a set of augmented capabilities within the browser, allowing the user to identify usability problems, suggest localized changes, or make quick modifications to the current page. The SO extension sends the changes made by the user to the SO server's repository of overlays. Upon a subsequent page load, the SO extension retrieves the overlays and requests made by all the community members associated with the current page from the server and applies these changes by modifying the DOM of the rendered web page.

This implementation allows us to gain complete control over the rendering of each web page. It also allows easy porting to extensions for other browsers or a bookmarklet application.

3 Study Design, Site, and Data Collection

After creating Social Overlays, we wanted to know whether it was a feasible approach. To evaluate the feasibility of Social Overlays, we conducted a three-phase study. In the first phase, we determined whether community members could use SO to identify and repair usability problems. We learned that people could use SO effectively, but we could not tell how good SO's results were compared to those generated by standard usability methods. Thus, in the next two phases we compared SO to two standard usability methods: expert inspection and usability testing. The results of all three phases show that SO is a feasible approach to website usability.

3.1 SO Evaluation

First, we needed to know whether community members without training in usability could identify and repair usability deficiencies on their website using SO. Therefore, in the initial phase of our study (which we will call the "SO evaluation"), we asked 13

[1] `http://developer.chrome.com/extensions/`

information science students, none of whom had any formal usability education or experience, to use SO in the process of completing four tasks on the website of their school (and ours), the University of X School of X (UXSX).

We did not choose the website and users as a matter of convenience. UXSX's website had been recently overhauled and had many flaws and problems, thus providing a good testbed. In addition, by using UXSX and its users, we could emulate website users coming to the webpages they use everyday and finding problems based on their normal tasks. Moreover, UXSX is an organization similar to those foreseen as our target communities. It consists of approximately 400 people, including students, staff, and faculty, and is reasonably cohesive socially.

We selected four common information–seeking tasks (see Appendix I) and asked participants to carry out the tasks, using SO to identify and, if possible, repair any issues they encountered as they went along. Each task had known usability issues associated with the website. While the four tasks selected covered only a small subset of the site's possible interactions, they were enough to observe the community process of discovering, noting, fixing and voting on changes to usability problems within a tractable timeframe. We anticipate that, if SO were deployed over a longer time period, a larger number of community members would visit many more pages, ultimately covering all of the most common interaction paths.

To emulate people coming to a web site over time, we ran our participants consecutively. Thus, the overlays (UI modifications) and requests made by a participant were available to all participants after him/her. Participants were first given a demonstration of using SO to fix four example usability problems, and then they were instructed to identify usability problems while solving the four tasks. The participants were asked to state their ideal solution to a problem, and then solve the issue using SO. If a problem could not be repaired by SO's Text, Link, and Tooltip tools, the participant had the option of submitting a request in SO or not doing anything. Each session lasted about an hour and each participant received a $10 coffee shop gift card.

3.2 Expert Inspection and Usability Test

As mentioned, to address whether SO worked as well as standard usability methods, we conducted two additional evaluations. In the second phase (which we will term the "expert inspection"), we asked four usability specialists with at least four years of professional experience to conduct an expert walkthrough (as described in [15]) of the same four tasks that were used in the SO evaluation. The experts were given the same demonstration of SO in the beginning of each session. They were asked to identify as many problems as they could while walking through the four tasks, and envision how SO could be used for implementing their suggested solutions. Each expert was compensated $50 for participating an hour-long session.

In the third and final phase (the usability test), we commissioned a team of external usability evaluators to conduct a conventional lab-based usability test on the UXSX site, and asked them to report the problems they found along with recommendations to address those problems. The usability team consisted of two graduate students with formal training in usability testing and one usability professional with formal training and three years of professional experience.

The usability test followed the standard protocol described in Rubin's widely-used textbook [16], which consists of pre–test and post–test questionnaires, task observation, and debriefing. Eight additional UXSX students participated in the test and tackled the same four tasks used in the SO evaluation and expert inspection. Each session in the usability test lasted about an hour and each participant received $20.

4 Evaluation Results

In this section, we answer the following questions:

- Within our evaluation study, could community members use SO to report a substantial number of usability problems on their website?
- In addition to identifying problems, were community members able to use SO to repair at least some of them in a helpful way?
- How well did the SO approach work, in comparison with expert inspection and usability testing?
- How did community members collaborate informally in using SO to improve their site and benefit from one another's efforts?

In short, is the SO approach likely to be viable?

4.1 Community–Based Usability Improvements

The results of our evaluation show that community members without training in usability can identify a large quantity of usability deficiencies on their website, at least in this community.

In our data analysis, we tallied all problems for which the SO evaluation participants either made an overlay or submitted a help request. We then manually verified these issues on the UXSX website, confirming the existence of identified problems and eliminating duplicates. As a group, they documented 47 unique problems in the process of solving the 4 evaluation tasks. These included issues that could and could not be fixed with SO, but excluded issues that were verbally reported only as well as issues that were similar to those used as examples in the SO demo. In the rest of this subsection, we describe the problems identified by community members in the SO evaluation, and how they as a collective addressed those problems.

Types and Characteristics of Overlays. Using the Text, Link, and Tooltip tools provided by SO, the participants in the SO evaluation made 50 overlays (i.e. page modifications) to address 27 (57.4%) of the total 47 problems they documented. Among the 50 overlays, 10 were alternative text or labels, 11 were hyperlinks attached to existing elements, and 29 were tooltips.

As expected, the Text tool was often used to correct or clarify a link's label. For example, P3 changed a link's label from "Course schedule" to "Course schedule by term," as she thought people might expect the linked schedule to be organized by week. Other uses of the Text tool included replacing an unfamiliar term, correcting typos, and appending a commonly-used acronym.

The Link tool was generally used to shorten navigational paths. For example, five different participants linked 5 static course titles on a degree requirement page to their respective course information pages, after they found it took too many clicks to check course information from the requirement page.

Usage of the Tooltip tool was more varied and interesting. Most of the tooltips were added to links in order to help users decide whether or not to click on them. First, tooltips were created for a link to hint what information could be expected in the linked page. For example, P7 attached the following tooltip to the “guest speakers” link on the Media page: “job candidate talks are accessible through this link.” Second, tooltips were used to clarify community-specific jargon. For example, P3 added the following tooltip to explain “faculty guest lectures” as “another way of saying ‘job talks’.” Lastly, P1 and P10 used tooltips to give specific directions to aid navigation. Fig. 3 shows a tooltip P10 created to direct users to other course schedule viewing options that were hard to find due to the poor information architecture.

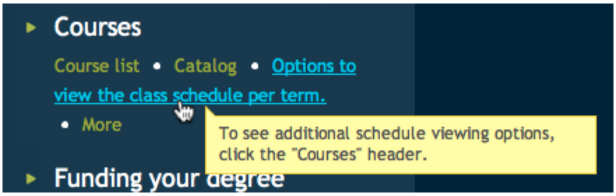


Fig. 3. P10 added a tooltip to indicate that there are other options to view the class schedule (e.g. by week) on the page linked to the “Courses” header.

Help Requests. Though the remaining 20 problems were not “overlaid” with page modifications, they were identified via 33 help requests from the SO evaluation participants. When did participants choose to submit a request rather than create an overlay? We found that 10 of those unresolved problems fell out of the scope of tools like SO. They were related to backend issues, missing content, or missing a feature. Among the remaining 10 unresolved problems, 3 of them could possibly be repaired with the current version of SO, while the other 7 issues could potentially become repairable using an improved version of SO outlined in the Discussion section.

We also examined those requests’ quality, placements on the page, and intended audiences. First, close to 90% of help requests included specific recommendations for improving the website, which suggests that participants generally had pretty clear ideas about how the problems they identified could be resolved.

Second, participants appear to be able to find an appropriate element to associate their requests to. Among the 33 requests submitted, 21 requests were attached to a specific page element, while 12 went into the “bucket” as page–wide issues or general comments. Most requests were attached to an appropriate element that helped us understand what the issues were.

Third and finally, many requests appeared to be added in the hope that webmasters would fix them from the server. However, there were also 4 requests intended to go to other community members, asking whether a change was accurate or dividing up the work of making a series of similar changes.

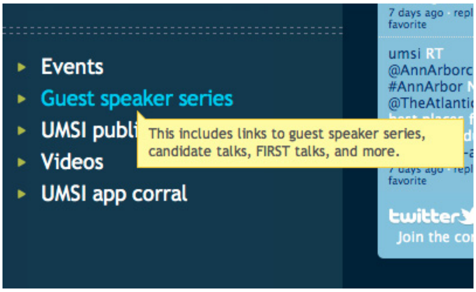


Fig. 4. P1 added a tooltip to the "Guest speaker series" link to indicate that users can find faculty candidate talks in the linked page. This tooltip significantly reduced the number of clicks taken by subsequent participants once they reached this page.

Quality of Community–Generated Usability Enhancements. Did the participants in the SO evaluation make helpful changes? Our analysis suggests that at least some user-generated overlays made subsequent participants more efficient in solving tasks. For example, one of these helpful overlays was P1's tooltip added to the "Guest speaker series" link on the "Events and News" page, as Fig. 4 shows. The tooltip explained what was in the linked page, which was linked to the faculty candidate talks that participants were asked to find in Task 1. The mouse click counts in Fig. 5 shows that Task 1 became substantially easier after P1 created that tooltip.

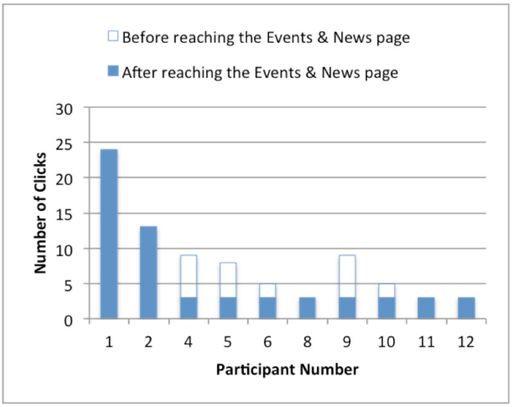


Fig. 5. The number of mouse clicks users made decreased dramatically after P1 added a tooltip to clarify a critical link on the Events & News page. Participants who did not take this route were omitted from the figure.

The participants made helpful overlays to simplify other tasks as well. For example, after completing task 3, P7 linked a course title to a page that provided additional information about the course, and then P12 came across it and said, *"So someone added a link. Oh man, it doesn't tell me if it has PEP credits or not. So someone* *helpfully probably put this link in. [Clicking the link]. That was helpful, that was totally helpful."*

However, not all overlays were well crafted. For example, there were tooltips added to clarify the difference between the course list and the course catalog, but apparently these tooltips still lacked detail that would help P12 pick one course page over another. Although the quality of the user–generated usability enhancements varied, they rarely made the website harder to use.

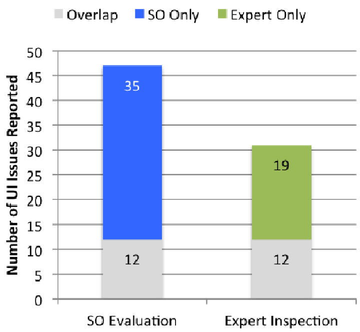
In summary, we found that Social Overlays could lead to a substantially improved website. However, this improvement was not uniform, and much of the improvement showed satisficing behavior.

4.2 Differences between Social Overlays and Standard Usability Methods

To assess and contextualize the problems identified by participants in the SO evaluation, we first compare them with the results of the expert inspection and then with those of the usability test. Specifically, we went through the list of issues reported in the SO evaluation and checked if each of them was covered by the expert inspection or the usability test.

To our surprise, participants in the SO evaluation documented 52% more problems than the 31 problems reported by the expert inspectors. Only 12 out of the 47 problems documented in the SO evaluation were also identified by the experts (see Fig. 6). How could the participants in the SO evaluation find these 35 additional problems? As members of the community, they appear to have advantages in three aspects, based on our analysis:

First, the community members in the SO evaluation leveraged their lived experiences in the organization during problem identification. For example, P3 pointed out that the Tracking and Planning Sheet, a useful PDF file that she could print out and use as a resource in completing task 4, was buried too deeply in the site. She was aware of this problem, because of her prior experience in planning courses as a master's student at UXSX.



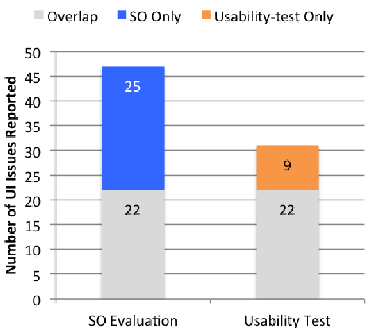


Fig. 6. The above bar charts show that community members using Social Overlays reported not only more but also different usability issues than the expert inspectors and the usability testing team. We describe those differences in section 4.2.

Second, community members uncovered similar problems that occurred in different places in the website, since as a whole, they were exposed to more pages and explored different paths than the group of experts.

Third, the community members were able to project the needs and preferences of the sub–community to which they belonged. For example, two masters' students, P4 and P11, believed that other students would want to find student club events in the school's official Events page.

Nevertheless, the SO evaluation participants did not identify 19 problems that were found in the expert inspection. It appears that the experts were able to spot these additional issues by following general principles and heuristics. For example, the experts were sensitive to inconsistencies on the website such as showing Prof. A's profile on the same page of a course taught by Prof. B. The experts also seemed to evaluate the site following best practices in usability. For example, E1, E2, and E3 mentioned that the website failed to visually separate groups of content on several pages.

Next, we compare the results of the SO evaluation with those of the usability test. The usability testing team reported 31 problems, some of which were broad statements and covered multiple narrower problems documented by the participants in the SO evaluation. In total, the usability test covered 22 of the 47 problems documented by the participants in the SO evaluation (see Fig. 6), representing a larger overlap than that between the SO evaluation and the expert inspection. This increased similarity is not surprising, since usability testing could also access community members' lived experience and local knowledge about the site.

But why did SO users report 25 problems that were not found by the usability testing team? From our analysis, there were two main reasons. First, participants showed empathy with their peers in the UXSX community, pointing out 12 issues that did not prevent themselves from solving tasks but they believed could be frustrating or confusing to less experienced peers. For example, P3 found it a potential problem that the acronym "PEP" was not linked to its full description page, though she already knew what PEP stood for ("Practical Engagement Program"). Second, participants in the SO evaluation pointed out 7 more issues related to missing information or features that they expected to see on the site.

In short, we found that community members reported more problems compared with usability experts and the external usability testing team, and that their reported problems differed in systematic but useful ways from the results generated by the expert inspection and the usability test.

4.3 Community Processes in SO

We also wanted to assess whether SO would show preliminary evidence of community activity. The most important community activity, as mentioned in section 4.1, was getting assistance from the webpage improvements made by other users. We also observed community members giving feedback on the overlays made by others. Six of our participants used the voting feature of SO, and they gave 15 thumbs–up and 6 thumbs–down to overlays made by prior participants. We did not observe conflicting votes on any of these changes. Other users verbally commented on the helpfulness of other users' overlays, but they did not click the voting button in SO. Four of the 6 thumbs–down were given to point out errors made by P8 and P9, while the other 2 thumbs–down were cast by P7 and P9, when they disliked existing overlays and sought to replace them with one of their own.

We also observed some evidence of social dysfunctionalities that might become more critical in larger communities or with more use. P4 showed self-serving tendencies, declaring, "*I would make that a link, but it would be motivated by my convenience and not out of altruism for other users.*" Another participant did not like all of the changes, but did not revert or modify those changes. (This might also be tacit approval of the change's sufficiency, however.) As well, participants did not vote as much as we hoped. We will return to these issues in the Discussion section.

4.4 Subjective Perception of the Utility of SO

Participants in the SO evaluation commented on what aspects of SO they found particularly helpful. First of all, participants liked the ability to immediately see the change they made using SO. Many of them were delightedly surprised when they saw problems repaired instantly. Second, participants liked that SO provides peripheral awareness of community activities through the indicators on the side margin (see Fig. 2), as P5 mentioned, "*I really like the scroll bar [indicators] that change colors. That's the first thing I looked to, besides the blinking.*" Third, participants appreciated that the changes they made were shared with other community members. P3 enthusiastically commented, "*It can only be in the long term a very big asset to the community, especially the social element ... where people vote up and down changes.*"

In summary, the above results of the SO evaluation and our comparisons with expert inspection and usability test show that:

1. With Social Overlays, community users identified a substantial number of problems as they interacted with the site. They were able to repair many of them by creating overlays helpful to others and address the rest by submitting constructive requests.
2. With Social Overlays, community users uncovered problems that existed along multiple paths of browsing and brought to the problem identification and solving process their lived experience and ideas that only members of the community would have.
3. With Social Overlays, community users seemed to take on a more active role in reporting UI problems than the participants in the usability test.

Our findings argue that Social Overlays is useful and feasible for communities, as least those similar to UXSX in size and cohesiveness, to collectively identify and address usability problems on their website. Moreover, the results of using SO were comparable to standard methods that are more expensive, such as expert inspection and usability testing, though each of the three approaches showed different strengths and priorities in our study.

5 Discussion

To summarize, this paper:

- Introduced a new community–based approach for addressing web usability problems without requiring expertise in usability evaluation or interface design.
- Presented a new system called Social Overlays that embodies this approach of collaborative usability improvement and demonstrated its technical feasibility.
- Provided an evaluation that showed that Social Overlays and the approach it enables are useful and feasible for an important set of communities and websites.

Despite SO's success, we found five issues with it and the approach in our evaluation. First, our evaluation showed that community users grounded problems in their lived experience inside their community, covered a larger number of web pages, and made quick responses to UI issues. This, we have argued, is critical for many organizations without access to usability professionals. However, we also saw that users using SO might need help formulating holistic and broad redesign recommendations. An interesting question we would like to explore in future work is how SO can be used with some lightweight expert involvement, e.g. with an expert synthesizing or gatekeeping modifications proposed by users.

Second, although our evaluation argued that SO was likely to be helpful for communities with a few dozen to several hundred members, there are many other types of communities with many more members, less social cohesion, and potentially less trust. Addressing these communities will require additional functionality. Some public-facing websites are subject to spam, problematic content, and attacks, requiring more security mechanisms. Larger or more public websites and communities may suffer more motivation issues, although they also potentially have more users for SO. All of these issues remain to be addressed in future work, and we plan to do so in subsequent efforts. Nonetheless, we were pleased that SO appears to be able to work in at least an important subset of small-to-medium communities. The same type of community exists in many educational settings, community organizations, charities, and the like. Nonetheless, we will not know all the issues until this tool is released into the wild and used by other sites and communities.

Third, the current version of SO provided only three simple modification tools. Even with these limited capabilities, our participants were able to address common and important usability problems with ease: confusing labels and vocabulary issues, broken and cumbersome navigational paths, and unclear site functionality. We believe, based on our evaluation, that additional tools would make SO even more helpful. We are currently investigating capabilities to insert new links into a page, convert the highlighted part of text into a hyperlink, and support the navigational use of tooltips with better hyperlinking, and address higher-level issues of page and site organization. We are also examining how to address potential information overload problems that would arise if SO were to scale, including new awareness visualizations and mechanisms for issue consolidation.

While it is technically possible to include more powerful and sophisticated tools such as a full-fledged style editor, we believe this would run the risk of having naïve users freeze or make too many mistakes. We also plan to examine this risk in future studies.

Fourth, based on even our limited study, we suspect there will be resulting problems as SO use grows in a site. As the number of users grows, it will increasingly be a concern that different subcommunities inhabit different language worlds, and individuals may wish to place markers for themselves that would be inappropriate for the entire community. We are currently investigating mechanisms to allow both subcommunities and individual personalization.

Finally, as mentioned, we had less usage of the voting feature than we had hoped. In our observations, non-voting behavior often implied agreement, and there were very few overlays or requests that were perceived negatively by participants. We expect both the use and the utility of the voting feature to increase as more users participate in an SO deployment; however, we are investigating new user interface designs to make voting easier and more meaningful.

We must also acknowledge the many limitations in the current study. Our tasks were artificial and scoped, as with any lab–based study. As well, the study was short–term and used selected pages. As mentioned, we will not know all the issues until SO is released into the wild and used by other communities—work that is presently ongoing. The current system and evaluation reported here, however, has led us to believe the usefulness and stability of SO are such that such explorations can fruitfully occur.

6 Related Work

The SO system and the approach it enables, as presented above, aim to tackle the challenge of post-deployment usability which is important but has been undervalued in the user-centered design practice [6, 19]. A number of commercial offerings have sprung up that facilitate the collection of user feedback on deployed websites, either through web forms (e.g., Feedback Army[2] and UserVoice[3]) or message boards (e.g., SuggestionBox[4]). Additionally, web annotation systems (e.g., Diigo[5] and AnnotateIt[6]) could potentially be employed to obtain user feedback, though most of them are focused on learning and information management. It is evident that, while being reasonable solutions for collecting user feedback, all of these tools require users to depend on website supporting staff to assess the severity of the problem, design a solution, and implement the change. In contrast, SO addresses usability by enabling a website's user community to not only identify problems but also repair many of them immediately.

To design a system like SO, we draw on three bodies of technical work: *web-rewriting systems*, *collaborative web accessibility tools*, and *community-based help systems*. We will describe each of them in turn.

---

[2] `http://www.feedbackarmy.com`
[3] `http://www.uservoice.com/`
[4] `http://www.suggestionbox.com/`
[5] `http://www.diigo.com/`
[6] `http://annotateit.org/`

First, we build upon *web-rewriting systems* that allow users to alter the design or behavior of a webpage at runtime. General-purpose web-rewriting systems like Chickenfoot [4] and GreaseMonkey [14] support such rewriting, and they have online repositories whereby users can share, discover, and make use of scripts created by others. However, the versatile capability of these systems creates difficulties for most users, since many users do not know how or want to inspect, edit, and debug scripts. SO leverages some of the same technologies used in web rewriting systems (e.g., JavaScript injection), but SO nonetheless has been designed specially to allow naive users to contribute.

Second, collaborative web accessibility tools (e.g. AccessMonkey [3] and Social Accessibility [18]), represent a specialized class of web-rewriting systems—one that shares SO's high-level goal to make websites easier to navigate and use. Accessibility systems, though, are focused on helping particular subsets of users who share particular disabilities. For example, the Social Accessibility system [18] allows volunteers to enter missing metadata that can then be consumed by visually impaired users who subsequently visit the augmented sites using screen readers. Though the underlying technologies that enable SO and Social Accessibility are similar, the sociality afforded by the SO approach is fundamentally different, because in SO both producers and consumers of collaborative usability enhancements belong to the same group. Our study has shown that many usability problems can best be identified and addressed by users who share lived experience relevant to the community of the site.

Therefore, the sociality supported by SO is closer to that studied in the third body of work we draw upon: *community-based help systems.* These systems assist users by providing information generated by other members of the community. Community-based help systems, such as Answer Garden [1], QuME [20], IP–QAT [9], and LemonAid [5], have focused on making it easier for users to ask questions and find answers, and these systems have shown resoundingly that users can collectively create useful information. Answer Garden builds a store of questions and answers, but accessing it requires using a separate system, resulting in a potential distraction from the core tasks users are trying to accomplish. IP-QAT and LemonAid associate questions with related UI elements, and display answers as part of the system for which help is sought. IP-QAT and LemonAid, however, do not seek to actually improve the user experience of the system by effecting design changes, as SO does, which has the additional benefit of reducing the need for seeking help in the future.

In short, none of the above systems are focused on post-deployment usability assessment and correction, though we draw inspirations and learn lessons from them to inform different aspects of SO design. By specifically designing for the collaborative work around post-deployment usability, Social Overlays provides new capabilities of community interaction and assistance.

7 Conclusions

In this paper, we presented the Social Overlays system (SO) and its evaluation. SO enables end–users to identify and repair common user interface problems on a website by making “overlays” that rewrite specific page elements. Moreover, SO displays page modifications to others who also have this extension installed.

We also reported findings from an evaluation of Social Overlays. In our study, we found that a group of community members without any usability training could use SO to identify and fix UI problems on their medium–sized community’s website. We also compared the results from those community members to the assessments of usability experts as well as the results of a usability test, and found that SO produced a larger number of issues, and that these issues partially overlapped with those found through standard methods. To summarize, we found that Social Overlays can be a viable approach. By having users improve a website’s usability as a part of their use of that site, a community can collectively make their website more usable.

Acknowledgements. This work was supported in part by the National Science Foundation (IIS 0905460), the University of Michigan Rackham Graduate School, and kynamatrix Research Network. We also thank Sonali Mishra, Pei-Yao Hung, and Chuan-Che Huang for their help with this study.

References

1. Ackerman, M.S.: Augmenting the organizational memory: a field study of answer garden. In: Proc. CSCW 1994, pp. 243–252. ACM, New York (1994)
2. Bias, R.G.: The pluralistic usability walkthrough: coordinated empathies. In: Nielsen, J., Mack, R.L. (eds.) Usability Inspection Methods, pp. 63–76. John Wiley & Sons, Inc., New York (1994)
3. Bigham, J.P., Ladner, R.E.: Accessmonkey: a collaborative scripting framework for web users and developers. In: Proc. W4A 2007, pp. 25–34. ACM, New York (2007)
4. Bolin, M., et al.: Automation and customization of rendered web pages. In: Proc. UIST 2005, pp. 163–172. ACM, New York (2005)
5. Chilana, P.K., et al.: LemonAid: selection-based crowdsourced contextual help for web applications. In: Proc. CHI 2012, pp. 1549–1558. ACM, New York (2012)
6. Chilana, P.K., et al.: Post-deployment usability: a survey of current practices. In: Proc. CHI 2011, pp. 2243–2246. ACM, New York (2011)
7. Hartson, H.R., Castillo, J.C.: Remote evaluation for post-deployment usability improvement. In: Proc. AVI 1998, pp. 22–29. ACM, New York (1998)
8. Hill, W.C., et al.: Edit wear and read wear. In: Proc. CHI 1992, pp. 3–9 (1992)
9. Matejka, J., et al.: IP-QAT: in-product questions, answers, & tips. In: Proc. UIST 2011, pp. 175–184. ACM, New York (2011)
10. McGrath, D., Spear, M.B.: The Academic Crisis of the Community College. SUNY Press (1991)
11. Muller, M.J., et al.: Methods & tools: participatory heuristic evaluation. Interactions 5(5), 13–18 (1998)
12. Nichols, D.M., et al.: Participatory usability: supporting proactive users. In: Proc. CHINZ 2003, pp. 63–68. ACM, New York (2003)
13. Nielsen, J.: Usability Engineering. Morgan Kaufmann Publishers Inc., San Francisco (1993)
14. McFarlane, N.: Fixing Web Sites with GreaseMonkey, `http://www.linuxjournal.com/article/8215?page=0,0`

15. Oppermann, R., Reiterer, H.: Software evaluation using the 9241 evaluator. Behav. Inf. Technol. 16(4-5), 232–245 (1997)
16. Rubin, J., Chisnell, D.: Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests. Wiley (2008)
17. Sproull, L., et al.: Connections: New Ways of Working in the Networked Organization. MIT Press (1992)
18. Takagi, H., et al.: Social accessibility: achieving accessibility through collaborative metadata authoring. In: Proc. ASSETS 2008, pp. 193–200. ACM, New York (2008)
19. Wagner, E.L., Piccoli, G.: Moving beyond user participation to achieve successful IS design. Commun. ACM 50(12), 51–55 (2007)
20. Zhang, J., et al.: QuME: a mechanism to support expertise finding in online help-seeking communities. In: Proc. UIST 2007, pp. 111–114. ACM, New York (2007)

Appendix I: Evaluation Tasks

1. Imagine you're a UXSX faculty member who has traveled a lot recently. As a result, you missed several faculty candidate talks. Now you want to find the video recordings of the following candidates' talks on the UXSX website: Jiang Chen, John Smyth, and Amit Gupta [names changed for publications]
2. Imagine you're a master's student at UXSX who stays in town this summer. You'd like to find out what's happening at UXSX during the summer.
3. Imagine that you're a new master's student. As part of the program requirements, you need to earn a certain number of PEP (Practical Engagement Program) credits. You're thinking about taking SX 622, so you'd like to find out how many PEP credits SX 622 offers.
4. Imagine you're a second–year master's student in the Human Computer Interaction (HCI) specialization. You're in the process of planning coursework for next semester. You'd like to find three HCI elective courses you're interested in and find out who is teaching each of these classes.

Usability Evaluation in a Digitally Emerging Country: A Survey Study

Fulvio Lizano[1], Maria Marta Sandoval[2],
Anders Bruun[1], and Jan Stage[1]

[1] Aalborg University, Department of Computer Science, Selma Lagerlöfs Vej 300, Aalborg East, Denmark
[2] National University, Informatics School, PO-Box 86-3000, Heredia, Costa Rica
{fulvio,bruun,jans}@cs.aau.dk, msandova@una.cr

Abstract. Several emerging countries experience increasing software development activities. With the purpose of provide useful feedback on possible courses of action for increasing application of usability evaluation in such countries, this paper explores the status of usability evaluation in a digitally emerging country. Our aim is to identifying common characteristics or behavioral patterns that could be compared with digitally advanced countries. We used an online survey answered by 26 software development organizations, which gave a snapshot of the application of usability evaluation in these organizations. We found many similarities with advanced countries, several completely new obstacles more connected with software development matters and a relatively positive improvement in the lack of "usability culture". These findings suggest good conditions to improve conduction of usability evaluations in digitally emerging countries.

Keywords: Usability evaluation, advantages, obstacles, digitally emerging countries.

1 Introduction

Usability evaluation is a relevant and strategic activity in software projects [6]. For the user, a high level of usability in a software system is important [8]. For the user organization, usability is important because it can provide benefits such as increased sales, productivity, lower training costs and reduced technical support for users [2].

Previous studies of the perception of usability evaluation have been focused on obstacles and advantages. Two specific cases have identified obstacles in software organizations. The first one was a survey in Northern Jutland, Denmark [2] (known in this paper as the "D-Study") and the second, which replicated the first, was made in Southern Italy [1] (known in this paper as the "I-Study"). The D-Study identified several obstacles to increased application of usability evaluation, e.g. developer mindset, resource demands, lack of understanding, customer participation, conducting tests and test participants [2]. In the case of the I-Study, major obstacles identified were resource demands, no suitable methods, developers' mindset and user availability.

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 298–305, 2013.

In addition, this study identified advantages of usability evaluation such as quality improvement, the users' satisfaction, resource saving and competitiveness [1].

These facts, together with other problems found in digitally advanced countries [3, 4, 5, 7, 9, 10, 11], can be visualized, as a whole, in a view formed by several dimensions of types of actors (e.g. users or clients, software developers, and organizations) plus types of facts (e.g. facts related to understanding, advantages and obstacles of usability evaluation) This is a way to put into context the real implications for usability evaluation in digitally advanced countries. In Table 1 we present the distribution of these findings into the cited dimensions (some references are related with different dimensions at once).

Table 1. Main findings related to usability evaluation

	Types of facts (related to…)		
Types of actors	Understanding of usability evaluation	Advantages of usability evaluation	Obstacles of usability evaluation
Users / Customers	– User involvement[1,2] – Customer involvement[2]	– User satisfaction[1]	– Test participants[2] – Customer participation[2] – User availability[1]
Software developers			- Developer mindset[2] - Developer mindset[1] - Lack of trained eng.usab/HCI[10] - Personal developers tools[11]
Organizations	- Functionality[2] - Problem/task solving[2] - Possibility test[2] - Usability evaluation[1] - Usability definition[1] - Accessibility test[1]	- Qlty. improvement[1] - Competitiveness[1] - Resource saving[1]	- Lack of understanding[2,10, 11] - Resource demands[2] - Conducting tests[2] - No suitable methods[1] - Resource demands[1, 10] - Resistance to UCD-Usability[10] - Lack of comms. of impact[10] - Lack coupling UCD & S-Dev l.cycle[11] - Gap SD & usability[11] - Edu. lack coupling (SD&usab.) [11] - Lack of respect and support[5] - Limited description HCI in SE[4] - Strong differences HCI & SE[7] - Lack of "usability culture" [3, 9]

In this paper, we present the results of a study that explored the application of usability evaluation in software development organizations in a digitally emerging country. The aim of our study was to explore the understanding of the usability evaluation concept, their advantages and obstacles. Our interest was identifying similarities, differences and patters that could enhance the application of usability evaluation in other digitally emerging countries. This explains why we will present our results and compare them to other studies, especially with the D-Study and the I-Study.

2 Method

2.1 Settings

We conducted a survey at software development companies in Costa Rica. According The Global Information Technology Report 2012 (World Economic Forum – www.weforum.org/gitr), Costa Rica is a digitally emerging country ranked in the 58 position of the Networked Readiness Index (NRI). Costa Rica has a NRI of 4 in a 1-to-7 scale.

2.2 Participants and Procedure

The study has involved companies that could potentially conduct usability evaluations and in addition met the specific criteria, e.g. located in a specific geographical area (Costa Rica), that develop software or hardware with graphical user interfaces, that develop software for customers or for internal use and that employ more than a single person. Initial set of participants was made using the list of organizations affiliated to the Chamber of Information and Communication Technology - CAMTIC, by its Spanish acronym (www.camtic.org) (148 organizations). This organization is open to any IT organization of Costa Rica and was founded in 1998. Because CAMTIC is open to a broad range of IT companies, we decided to filter the original list obtaining a final list of 35 organizations. Our survey was completed by 26 organizations (74%). The average number of years of operation of the organizations participating in the study was 11. The average of age of the persons, who filled the questionnaire, was 39. 15% of them were females. All these companies were located at the central valley of Costa Rica (the most developed zone of Costa Rica). The organizations had this distribution on number of employees: 58% (1-10), 19% (11-50) and 23% (51-250). In order to find the person most appropriate to participate in the study, we contacted every company personally by phone in order to enquire who could provide an opinion that could reflect the position of the company in the survey. These persons received an electronic token to access the online survey which was active for 4 weeks.

2.3 Data Collection and Analysis

The questionnaire used in the study contained several parts. The main parts were: demographic and general information, products/services provided by the organization, methodology used to develop software applications, understanding of the usability evaluation concept, and obstacles and advantages of the application of usability evaluation. We used a combination of open-closed questions was used. The aim of open questions was to permit participants to express themselves in their own words. Closed questions were used in order to allow to them to reconfirm data previously provided. We used different analytical approaches to analyze the data generated in the closed and open questions. A quantitative analysis was used on the closed questions, and grounded theory by Strauss and Corbin [12] was used for the analysis of the open questions.

3 Results

We were interested in exploring the understanding of the concept of usability evaluation in the software development organizations. Our results allowed us to identify several categories of understanding that the organization have about this concept. In order to verifying what the organizations had understood by usability evaluation, after provide their definition of usability evaluation, we showed them a definition of usability evaluation based on the ISO-9241 standard. Next, we asked them if they made usability evaluations in their companies in accordance with this ISO definition and the strategy followed by them to do it. The participants basically reported two categories of strategy: internal or external conduction of usability evaluation. However, a relevant number of participants reported do not conduct usability evaluations at all. Finally, some participants did not provide a response for this enquiry. In Table 2, we present these results.

Table 2. Distribution of the strategy used to conduct usability evaluation related to the understanding of the concept of usability evaluation

	Category of understanding (In terms of…)				
Strategy used to conduct usability evaluation	Usability concept	Usability evaluation concept	Another kind of testing	No response	Percent
Internally	4	8	4		62%
Externally		1	2		12%
No Usab.evaluation	1	2	2		19%
No response				2	8%
Percent	19%	42%	31%	8%	

On the other hand, following the methodological approach established in our method, we identified the main advantages and obstacles for applying usability evaluation. First, using an open question we identified advantages and obstacles of the application of usability evaluation. Some participants offered more than one advantage/obstacle and others did not respond. The results were grouped in different categories of advantages/obstacles. In the case of the advantages, main results were product quality (35%), user acceptance (32%) and no advantages (32%). The main obstacles detected were users (22%), software design (19%), software development method (15%), costs (11%), software developers (4%) and no obstacles (30%). The organizations which reported do not conduct usability evaluation in their process, were the ones which had not found advantages or obstacles.

In order to complement the previous results, we presented to the participants with a list of common advantages and obstacles of usability evaluation. The participants could select more than one option. Results are showed in Table 3.

Table 3. Advantages and obstacles provided through closed questions

Advantages.	#	P	Obstacles	#	P
User satisfaction	19	39%	Recruitment of test participants	9	20%
Quality improvement	18	37%	Conduct test / no method	6	13%
Competitiveness	5	10%	Developer mindset	17	38%
Resource saving	6	12%	Resources demands	8	18%
Other	1	2%	Other	5	11%
Total	49			45	

These results represent some of the main facts related to usability evaluation in digitally emerging countries. To facilitate a comparison of these facts with the ones presented in digitally advanced countries, we present them in a similar way as we did at table 1. Thus, as it is possible to see in Table 4, users, customers and software developers are not presented into the understanding of usability evaluation, by the software development organizations at the digitally emerging countries. In addition, our results suggest a weak visualization of advantages and obstacles in the same context. Similar to the digitally advanced countries, at digitally emerging countries it is possible to find more facts related to the understanding, advantages and obstacles of usability evaluation, in the context of the organizations.

Table 4. Summary of results

	Types de facts (related to…)		
Types of actors	Understanding of usability evaluation	Advantages of usability evaluation	Obstacles of usability evaluation
Users / Customers		– User acceptance – User satisfaction	– User – Test participants
Soft.Dev.			– Software developers mindset
Organizations	– Usability concept – Usability evaluation concept – Other test	– Product quality – Qlty improvement – Competitiveness – Resource saving	– Software design – Software development method – Costs – Usability evaluation conduction. – No usability evaluation method

4 Discussion

Our results suggest a relatively good understanding of the understanding of the usability evaluation concept, including some similarities to previous studies, e.g. in some aspects of the notion of usability [1], and in usability matters as a whole, specifically in some responses related to user involvement [2]. The good understanding about the definition of usability evaluation was obtained from organizations which conduct usability evaluation internally. This practical experience supported this better understanding. This is even more evident when we analyze the reasons given by those organizations that do not conduct usability evaluations (19%). Although the distribution of their understanding is uniform in the different categories of understanding, these participants provided opinions that are clear signals of a misunderstanding about usability evaluation, e.g. “in open source software projects you do not need usability evaluations” or “some projects do not require usability evaluations” and “a software project only needs functional tests”. Here, we can see and excellent example of what the lack of "usability culture" is [3, 9].

In the case of advantages, our results are fully in agreement with the I-Study [1]. However, in the case of the obstacles, our study found very interesting results. The ‘user’ was identified as one of the most relevant obstacle. This was emphasized by participants who conduct usability evaluations internally, which makes this result conclusive. Both the D-Study and I-Study also identified this obstacle but with a lower level of importance [1, 2]. This finding allows us to notice that in a digitally emerging country, participation of users in usability evaluation seems to be particularly challenging. Consistently with the D-Study and the I-Study, our study confirmed obstacles related to resource demands and software developers’ mindset [1, 2]. It is interesting to notice that the level of relevance given to this last obstacle has changed across the D-Study, the I-Study and our study. This obstacle was very important in the D-Study. In the I-Study, its relevance was lower. Finally, in our study this obstacle is the last one mentioned by the participants. This change could, initially, reinforce our perception of a positive change in the lack of a "usability culture”. However, other results obtained in the closed questions seem to offers contradictory results. In this case, the most important obstacle selected by participants was related to the software developers’ mindset problem. These different levels of relevance are not necessarily a contradiction. Actually, the fact that this obstacle was cited twice in our study allow us to conclude that this matter continues been one of the most recognized obstacles against increased use of usability evaluation. The second obstacle identified in this part of our study is related to resource demands, which is not surprising.

In addition, there are new obstacles that were identified in our study. First obstacle is related to problems in the design process of the software, which subsequently could hinder conduction of usability evaluations. Second, a new obstacle was identified in some problems related to the software development method. This obstacle was identified by the I-Study in 2011 but not by the D-Study in 2008. Here it is possible to observe a change of tendency in the lack of "usability culture", into those organizations that have practical experience; an alternative view about the new obstacles, which is more connected to the software development process, seems to emerge to reduce some problems such as the confusion, the lack of coupling and some gaps between SE and HCI [11].

In digitally advanced countries the main facts related to understanding, advantages and obstacles to conduct usability evaluation are more connected with methodology and the organization (see Table 1). Users, customers and software developers have a lower visualization. More remarkable is the fact that software developers are not presented at all in such dimensions. Only in the case of the obstacles, it is possible to find more facts related to users, customers and developers.

In digitally emerging countries, this situation seems no to be better (see Table 4). Into the understanding of usability evaluation, the users, the clients and the developers are excluded at all. Only a limited number of advantages were noticed for users and clients, none related to developers.

We think that our study provide interesting results that can be extended to other similar contexts. The digitalization level and other human and economical indicators are pretty similar to others countries in the same region, e.g. Ecuador, Trinidad & Tobago, Panama, Peru, Brazil and The Bolivarian Republic of Venezuela. The average on Networked Readiness Index (NRI) in these countries (including Costa Rica) is 3,71 (SD=0,32), the mean value for the GNI per capita in PPP terms (constant 2005 international $) is 12,051 (SD= 4,740), the average of expected years of schooling is 9,05 (SD=2,21), the mean value for expectancy of life (years) is 74,81 (SD= 2,79). Main differences are related to population and territorial extension. Considering all these facts, the context studied in our research can be considered a good referent about how usability evaluation is conducted in other digitally emerging countries.

5 Conclusion and Future Works

In this research we have explored the application of usability evaluation in software development organizations in the digitally emerging countries. To accomplish this, we conducted a questionnaire survey with 26 participating software development organizations. As part of our research, our findings were contrasted with results from similar studies in digitally advanced countries. The aim of our study was to obtain valuable feedback that could orientate future enhancement actions of application of usability evaluation in digitally emerging countries.

Our study found a relatively acceptable conduction of usability evaluation in digitally emerging countries, embodied by a fairly clear understanding about the meaning of usability evaluation and similar advantages and obstacles to the found in other digitally advanced countries. In addition, our research has identified new obstacles such as the users' behavior and problems related to the design of the software. These new obstacles can offer to HCI theory a complementary perspective on usability evaluation. These new findings seem to imply a decreasing tendency in the lack of "usability culture". However, our results do not permit strong conclusions about this matter as it was not a focus of our study.

However, any improvement of conduction of usability evaluation at the context studied must necessarily go through an empowerment process of users, clients and software developers, as main actors in such processes. In the case of users and clients, reasons to do that are more than evident; in some sense, these actors are a main cornerstone of theory of HCI. For developers, this strategy should help continuing improvement of some well studied problems, e.g. confusion, the lack of coupling and the gaps between software engineering and HCI. Future works could focus on exploring specific forms to enhance and increase the use of usability evaluations in software development organizations located in digitally emerging countries.

Acknowledgments. The research behind this paper was partly financed by Universidad Nacional (Costa Rica), MICIT (Costa Rica), CONICIT (Costa Rica), and the Danish Research Councils (grant number 09-065143).

References

1. Ardito, C., Buono, P., Caivano, D., Costabile, M.F., Lanzilotti, R., Bruun, A., Stage, J.: Usability Evaluation: a survey of software development organizations. In: Proceedings of 33rd International Conference on Software Engineering & Knowledge Engineering, Miami, FL, USA (2011)
2. Bak, J.O., Nguten, K., Risgaard, P., Stage, J.: Obstacles to Usability Evaluation in Practice: A Survey of Software Development Organizations. In: Proceedings of the 5th Nordic Conference on Human-computer Interaction: Building Bridges, pp. 23–32. ACM, New York (2008)
3. BugHuntress QA Lab: Mobile usability testing Problems and solutions. Research Report, Quality Assurance: Management & Technologies (QAMT), Ukraine (2007)
4. Ferre, X., Juristo, N., Moreno, A.M.: Obstacles for the Integration of HCI Practices into Software Engineering Development Processes. In: Ghaoui, C., Idea Group Reference (eds.) Encyclopedia of Human Computer Interaction, pp. 422–428 (2006)
5. Gulliksen, J., Boivie, I., Persson, J., Hektor, A., Herulf, L.: Making a difference: a survey of the usability profession in Sweden. In: Proceedings of the Third Nordic Conference on Human-computer Interaction (NordiCHI 2004), pp. 207–215. ACM, New York (2004)
6. IEEE Computer Society: SWEBOK Guide to the Software Engineering Body of Knowledge (2004), `http://www.computer.org/portal/web/swebok`
7. Juristo, N., Ferre, X.: How to integrate usability into the software development process. In: Proceedings of the 28th International Conference on Software Engineering (ICSE 2006), pp. 1079–1080. ACM, New York (2006)
8. Lindgaard, G., Chattratichart, J.: Usability Testing: What Have We Overlooked? In: CHI 2007 Proceedings, Usability Evaluation, San Jose, CA, USA (2007)
9. Muzaffar, A., Azam, F., Anwar, H., Khan, A.: Usability Aspects in Pervasive Computing: Needs and Challenges. International Journal of Computer Applications 32(10), 18–24 (2011)
10. Rosenbaum, S., Rohn, J.A., Humburg, J.: A toolkit for strategic usability: results from workshops, panels, and surveys. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2000), pp. 337–344. ACM, New York (2000)
11. Seffah, A., Metzker, E.: The obstacles and myths of usability and software engineering. Commun. ACM 47, 71–76 (2004)
12. Strauss, A., Corbin, J.: Basics of qualitative research. Techniques and Procedures for Developing Grounded Theory, 2nd edn. SAGE Publications (1998)

Understanding Diversity – The Impact of Personality on Technology Acceptance

Kay Behrenbruch[1], Matthias Söllner[2], Jan Marco Leimeister[2], and Ludger Schmidt[1]

[1] Kassel University, Human-Machine Systems Engineering, Mönchebergstr. 7, 34125 Kassel, Germany
[2] Kassel University, Information Systems, Pfannkuchstr. 1, 34121 Kassel, Germany
{K.Behrenbruch,Soellner,Leimeister,L.Schmidt}@uni-kassel.de

Abstract. Technology is becoming increasingly automated, aiming to ease the life of its users. However, besides the advantages of this trend, users are also faced with increasing risks, e.g., regarding their privacy. Examples are seamless online payments that come with the requirement to provide sensitive, e.g., credit card information, or social networks trying to elicit private information for its users. Research on technology acceptance identified two important factors for the individual decision to accept such kinds of risk: trust and personality traits. In this paper we present a model that integrates research findings for personality traits and for trust in the context of technology acceptance. We show that specific personality traits have a distinct direct or moderating effect. We, e.g., found that two personality traits moderate the relationship between perceived ease of use and intention to use. This moderation could explain the inconsistent findings on this relationship in prior research.

Keywords: Technology acceptance, personality traits, trust.

1 Introduction

On the one hand, many people publish and discuss detailed data of their lives on internet platforms like “quantifiedself.com”, and millions of people share their daily activity in online social networks. On the other hand there are others who feel uncomfortable with using their smartphones, because they want to protect data concerning their movement profile. Again others avoid e-commerce or e-banking to protect sensitive data. Some people are trying many new apps on their smartphones that seem in some way useful to them. Many others avoid new technology because they fear not to be able to operate it or even to damage it. Summing up, there is a great diversity in reactions to new technology in terms of acceptance. For the developers of new technology it is of high interest to understand those differences to be able to address them in their design. Ideally research should provide reliable information on the topic in the form of types of users with different characteristics that are matched with concrete needs for specific design features. The base of such an approach is to build a user model that combines characteristics of the potential users with their tendency to accept or reject new technology. This model needs to be scrutinised and checked to

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 306–313, 2013.

provide trustworthy information and to avoid undue stigmatisation of the users. The presented work contributes to improve and extend models that consider the influence of personality on acceptance.

In the beginning, research on technology acceptance concentrated on new occupational software in organizations and institutions [1]. Research presented in this paper focuses on web-enabled software in private use, with private data being processed and stored. In our paper, we focus on the influence of personality traits on technology acceptance. Many new products are not only unfamiliar or evoke resistance just because things are different than before. Furthermore, there is an increasing risk to lose privacy or money. That is why we pick up approaches that integrate trust in models for technology acceptance [2].

After presenting related work we develop hypotheses and a research model for the influence of personality traits on technology acceptance. We present results on basis of an empirical study with a social software that is open for integrating adaptive services for smartphones and discuss those results.

2 Related Work

The effect of personality on technology acceptance has been demonstrated by various authors. Especially, the use of social networks has been a major focus of research [3]. Another field of attention has been personality in acceptance of E-commerce [4]. Many aspects of personality were considered in HCI-research such as the need for recognition and sociability [5], computer anxiety, innovativeness, computer self-efficacy, and resistance to change [6], [7], as well as interest in the application domain [8]. The most used characterization of personality in this context is the five factor model FFM [9]. FFM encompasses five personality traits: extraversion, agreeableness, conscientiousness, openness to new experience, and emotional stability [10].

The most used and tested model for technology acceptance is the technology acceptance model TAM [1]. Within this model two beliefs have causal effects on the intention to use technology: perceived ease of use (PEOU) and perceived usefulness (PU). For the case of applications that put their users to some kind of financial or privacy risk, Gefen et al. [2] expanded TAM by another belief that affects the intention to use, namely trust.

3 Model Development

Devaraj et al. [9] developed and tested a research model that integrates TAM with personality traits for collaborative technology. This model incorporates FFM and computer self-efficacy. In this paper we adapt and extend this approach for a new class of technology: web-enabled software in private use, with private data being processed and stored. We integrate trust as an important believe influencing intention to use in accordance with authors [2]. Also we integrate the renowned concept of computer anxiety [11] as an additional personality trait. The relationship of the two concepts computer self-efficacy and computer anxiety is controversial in the literature.

Some authors argue for causal influence of computer anxiety on computer self-efficacy [12] others the other way round [13]. Our research model concentrates on the causal relationships between the three believes trust, PEOU and PU with intention to use and abstracts from the interdependency of the three beliefs specified in prior literature [1], [2], [9]. Some authors discuss a direct influence of personality traits on Intention to use [14], [15]. In accordance with the theory of reasoned action [16], that was used in the development of TAM, we hypothesize an influence of personality traits on the beliefs that determine the intention to use. Corresponding to our argumentation we hypothesize direct influence of personality on the beliefs or moderating effect on the causal relationship between beliefs and intention to use. Concerning trust it has to be kept in mind, that our research model is not addressing the influence of personality traits on trust in general, but on trust in a special kind of technology.

3.1 Hypothesis and Research Model

Below we develop hypotheses and a structural model. We did not find equally strong and plausible arguments for an influence of conscientiousness and agreeableness on the integrated beliefs as for the other three concepts of FFM. For example Svendsen et al. [8] do not integrate agreeableness in their model and restrict the hypothesized influence of conscientiousness on acceptance to a subset of the user group. Thus, we did not integrate conscientiousness and agreeableness into our research model.

Emotional Stability. A person who is emotionally stable would remain calm in many situations and would feel secure even when risks occur. An emotionally instable person shows the tendency to experience vulnerability. Feeling secure supports trust while feeling vulnerable makes it much more difficult to bring oneself to build up trust. Zhou and Lu [4] argue that a person with low emotional stability feels a lack of control and will not readily build trust in e-commerce.

H1. Emotional Stability will be positively associated with trust.

Extraversion. The surgency of extraverted people supports a trusting attitude because there is a lot of energy and readiness to overcome doubts and use technology for social interaction. Persons with low extraversion show the tendency to be more reserved and bringing about little motivation for building up trust for new technology and to open up ways for new kinds of social interaction. Zhou and Lu [4] argue that mobile services are a possibility for extroverted people to communicate with their friends at anytime from anywhere. Thus they will more readily give trust.

H2. Extraversion will be positively associated with trust.

Innovative software, especially on smartphones, has a high potential to foster social interaction. Extroverted people who are talkative and keen on social interaction with others will see much more benefit and use in getting in touch with such kind of technology than introverted ones who are happy with being on their own. Rosen and Kluemper [17] are in line with this argumentation: “Social networking sites would be another way for extroverts to assert themselves and thus could be considered quite useful (p. 4).”

H3. Extraversion will be positively associated with PU.

Openness. Interest in novelty fosters the willingness to build up in intention to use new technology on bases of a certain level of PU. People who prefer familiarity over novelty will possibly reject to use the technology with the same level of PU. “People who exhibit this personality trait [openness] seek out new opportunities to exhibit their creativity, and social networking websites are one way to do so. Also since these individuals are more intellectually curious than their peers, a novel way to communicate with friends and associates should be appealing to these people (p. 4).” [17].

H4. Openness will moderate the relationship between PU and intention to use such that the relationship is stronger for individuals with higher openness.

Closed people prefer the simple, plain and straightforward things. Using new technology is not easy right from the start. With lacking interest in the variety of new software a certain level of PEOU won’t be enough to build up an intention to use the technology. Open people like to spend time reflecting on things, they enjoy unusual interaction and are willing to overcome some possible challenges by learning when using new technology. Svendsen et al. [8] argue for a positive relation of openness to PEOU: “Openness to experience is an obvious candidate as this personality trait characterises an individual’s approach to new situations (p. 5).”

H5. Openness will moderate the relationship between PEOU and intention to use such that the relationship is stronger for individuals with higher openness.

Computer Self-efficacy. People who are confident in interacting with computers and information technology are in a better position to get along with a low level of trust concerning new applications. They probably will feel the ability to compensate the perceived lack of safety. When people have doubts about their capability to use such systems a small risk will probably prevent them from building up on the intention to use this system.

H6. Computer self-efficacy will moderate the relationship between trust and intention to use such that the relationship is stronger for individuals with higher Computer self-efficacy.

Furthermore, related research findings show that users anchor their perception of ease of use to computer self-efficacy [18]. Computer self-efficacy has a major impact on an individual’s expectations towards using computers. Contrary to Venkatesh and Bala [18] we do not expect computer self-efficacy to influence PEOU directly, but to moderate the effect of PEOU on intention to use. The belief that one will be able to perform a task with the system in question will alter the willingness to use a system at a given level of PEOU in the first place. From our point of view computer self-efficacy is not a major factor for PEOU. The reason for this is that individuals with high computer self-efficacy will feel to be able to deal with expected problems rather than expecting not to have problems at all.

H7. Computer self-efficacy will moderate the relationship between PEOU and intention to use such that the relationship is stronger for individuals with higher computer self-efficacy.

Computer Anxiety. The anxiety about the implications of computer use such as the loss of important data or fear of other possible mistakes, like damaging something prevent people from deciding to use a technology [11]. This factor is important for the level of the PEOU. With software that seems very easy, there may be a chance to overcome computer anxiety. When things seem very complicated the probability of rejection is very high. People with a low computer anxiety should be much more likely to build up an intention to use with different levels of PEOU. Nov and Ye [7] postulate that computer anxiety will be negatively related to PEOU of a digital library. The authors do not bring forward arguments for this notion. We prefer to hypothesize a moderating effect, because we claim that the perception of the usability of the system is not altered by anxiety: It is the assessment of the PEOU in comparison to fears of doing something wrong that is the expectable effect.

H8. Computer anxiety will moderate the relationship between PEOU and intention to use such that the relationship is stronger for individuals with lower computer anxiety.

To some people interacting with software is a threat. They fear to do something wrong. The higher perceived risk of something bad to happen will diminish the tendency to trust the software.

H9. Computer anxiety will be negatively associated with trust.

4 Methodology

We used a laboratory experiment with 344 undergraduate business students. The participants used a mobile application that was developed within a multi-disciplinary research project and supports mobile social networks, group interaction and mobility. It was designed for the group of 20- to 35-year-olds. The application supports its users in organizing meetings with friends at events. Events can be public or private happenings like concerts or parties. The application helps to plan the event, navigate the user to the event, and supports the user at the event itself. It proposes events to the users that fit their individual interests as well as their personal calendars. The application suggests means of transport and informs about possible delays, e.g. when using public transport. At the event, information concerning the location, the program, and tickets can be retrieved from other services that can be integrated. If desired by the user the application is able to suggest new friends based on the user's settings and interests. Within the laboratory experiment, the students received information on the idea of the system, how it works, and how to interact with the application. Afterwards, all students were asked to complete four predefined tasks using the mobile application, ensuring that participants recognized all functionalities of the system. Then all participants were asked to fill out a questionnaire, including items for FFM [19], Computer Anxiety [11], Computer Self-Efficacy [13], Trust [22], PEOU and PU [20]. Responses were recorded on a bipolar 9-point Likert response format. To achieve high quality results, we implemented several reverse coded items into the questionnaire, and checked all cases regarding the consistence of the answers given to the items relevant for our data analysis and the reverse coded control items. Furthermore, we dropped all cases that included missing data to ensure high data quality. In the end, 272 data sets were included in our analysis. We relied on SPSS 20 and SmartPLS 2.0 [23] to analyse our data. We used the approach of Chin [24] to model and evaluate the moderating effects in PLS.

5 Results

Before discussing the results regarding our hypotheses, we first need to assess the quality of our reflective measurement models. Consequently, we check the average variance extracted (AVE), the composite reliability, the indicator loadings, cross-loadings and correlations among our constructs as quality criteria [24]. Summing up, all values were above the required quality criteria. The lowest AVE was observed for PEOU (0.7640 > 0.5). The lowest composite reliability was observed for trust (0.9266 > 0.6) and the lowest indicator loading was 0.868 (> 0.7). Regarding the cross-loadings, all indicators showed the highest loading on their desired construct and the AVE for each construct was higher than any correlation with another construct. Consequently, our measurement models are reflective and reliable, and we thus can now confidently turn towards the evaluation of our hypotheses.

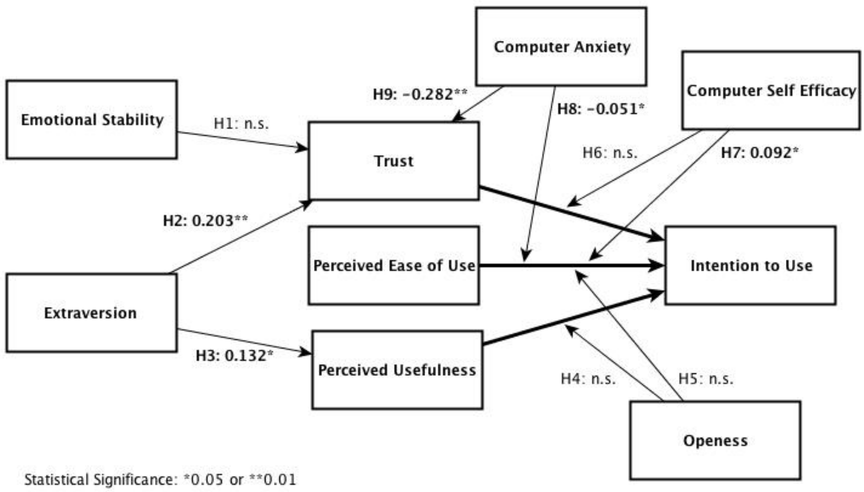


Fig. 1. Research model with results for hypotheses H1 to H9

Overall, we found support for five of our nine hypotheses. We could show that there is a significant impact of extraversion on trust (H2, path coefficient 0.203, $p < 0.01$) and PU (H3, 0.132, $p < 0.05$). Furthermore, we could show that there is a significant negative impact of computer anxiety on trust (H9, -0.282, $p < 0.001$). Additionally, we could show a negative moderation effect between computer anxiety and PEOU (H8, -0.051, $p < 0.05$) and a positive effect between computer self-efficacy and PEOU (H7, 0.092, $p < 0.05$). We could not find a significant impact of emotional stability on trust (H1, -0.075, not significant). Furthermore, we did not find empirical support for three of our supposed moderation effects. In particular the moderation effects between openness and PEOU (H4, 0.002, n.s.), between openness and PU (H5, -0.002, n.s.), and between computer self-efficacy and trust (H6, -0.010, n.s.). A graphical illustration of the results regarding our hypotheses is presented in Fig. 1.

6 Discussion

With respect to the general personality traits of FFM our results support the elevated position of extraversion with respect to technology acceptance. Both hypotheses for extraversion could be validated in our study.

Concerning the technology related item computer anxiety we found a new and significant relationship with trust (H9). The approved moderating effect of computer anxiety and computer self-efficacy on the causal relationship of PEOU and intention to use (H7 and H8) contradicts with the formation of hypotheses and the findings of Venkatesh and Bala [18]. In accordance with Venkatesh and Bala [18] we did not expect an influence of the two technology related items on PU. Our findings concerning PEOU can contribute to another current research topic: Gefen and Straub [21] ask for the reason of the changing importance of PEOU for intention to use in different studies and state that this varying importance of PEOU may be related to the nature of the task. Derived from our research results also the user group might be important with respect to the expression of characteristics for computer anxiety and computer self-efficacy. Our ongoing research will focus on verifying our findings for diverse user groups. The findings concerning relationship of PEOU and personality traits (H7, H8) needs to be attuned and further developed with respect to other research findings [18], [21]. In this context the following concepts will be also considered: personal innovativeness [14], computer playfulness and perception of external control [18].

Acknowledgment. We thank Hesse's Ministry of Higher Education, Research, and the Arts for funding our work as part of the research funding program LOEWE.

References

1. Davis, F.D., Bagozzi, R.P., Warshaw, P.R.: User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. Management Science 35(8), 982–1003 (1989)
2. Gefen, D., Karahanna, E., Straub, D.W.: Trust and TAM in online shopping: an integrated model. MIS Q. 27(1), 51–90 (2003)
3. Amichai-Hamburger, Y., Vinitzky, G.: Social network use and personality. Computers in Human Behavior 26(6), 1289–1295 (2010)
4. Zhou, T., Lu, Y.: The Effects of Personality Traits on User Acceptance of Mobile Commerce. International Journal of Human-Computer Interaction 27(6), 545–561 (2011)
5. Hughes, D.J., Rowe, M., Batey, M., Lee, A.: A tale of two sites: Twitter vs. Facebook and the personality predictors of social media usage. Computers in Human Behavior (December 2011)

6. Nov, O., Ye, C.: Personality and Technology Acceptance: Personal Innovativeness in IT, Openness and Resistance to Change. In: Proceedings of the 41st Annual Hawaii International Conference on System Sciences, HICSS 2008, p. 448 (2008)
7. Nov, O., Ye, C.: Users' personality and PEOU of digital libraries: The case for resistance to change. Journal of the American Society for Information 59, 845–851 (2008)
8. Svendsen, G.B., Johnsen, J.-A.K., Almås-Sørensen, L., Vittersø, J.: Personality and technology acceptance: the influence of personality factors on the core constructs of the Technology Acceptance Model. Behaviour & Information Technology, 1–12 (February/March 2011)
9. Devaraj, S., Easley, R.F., Crant, J.M.: How Does Personality Matter? Relating the Five-Factor Model to Technology Acceptance and Use. Information Systems Research 19(1), 93–105 (2008)
10. McAdams, D.P.: Five factor model in personality. Journal of Personality 60(2), 329–361 (1992)
11. Bozionelos, N.: Computer anxiety: relationship with computer experience and prevalence. Computers in Human Behavior 17(2), 213–224 (2001)
12. Bennett, J., Perrewé, P.L.: An Empirical Examination of Individual Traits as Antecedents to Computer Anxiety and Computer Self-Efficacy. Management Information Systems 26(4), 381–396 (2012)
13. Compeau, D.R., Higgins, C.A.: Computer Self-Efficacy: Development of a Measure and Initial Test. MIS Quarterly 19(2), 189–211 (1995)
14. Sim, J.J.: Influence of Personal Innovativeness, Self-efficacy and Subjective Norm in M-commerce Acceptance?: A Conceptual Review. International Journal of Network and Mobile Technologies 3(3), 40–48 (2012)
15. Amichai-Hamburger, Y.: Internet and personality. Computers in Human Behavior 18(1), 1–10 (2002)
16. Fishbein, M., Ajzen, I.: Belief, Attitude, Intention and Behaviour: An Introduction to Theory and Research. Addison-Wesley (1975)
17. Rosen, P.A., Kluemper, D.H.: The Impact of the Big Five Personality Traits on the Acceptance of Social Networking Website. In: Proceedings of the Fourteenth Americas Conference on Information Systems (2008)
18. Venkatesh, V., Bala, H.: Technology Acceptance Model 3 and a Research Agenda on Interventions. Decision Science 39(2), 273–315 (2008)
19. Gosling, S.D., Rentfrow, P.J., Swann, W.B.: A very brief measure of the Big-Five personality domains. Journal of Research in Personality 37(6), 504–528 (2003)
20. Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D.: User Acceptance of Information Technology: Toward a Unified View. Management Information Systems 27(3), 425–478 (2010)
21. Gefen, D., Straub, D.W.: The Relative Importance of PEOU in IS Adoption: A Study of E-Commerce Adoption. Journal of the Association for Information Systems 1(1), 8 (2000)
22. Söllner, M., Hoffmann, A., Hoffmann, H., Wacker, A., Leimeister, J.M.: Understanding the Formation of Trust in IT Artifacts. In: Proceedings of the International Conference on Information Systems (ICIS), Orlando, Florida, USA (2012)
23. Ringle, C.M., Wende, S., Will, A.: SmartPLS 2.0. Hamburg (2005)
24. Chin, W.W.: The Partial Least Squares Approach to Structural Equation Modeling. In: Marcoulides, G. (ed.) Modern Methods for Business Research. LEA, London (1998)

A Vocabulary to Access Users' Cultural Perspectives in Human-Computer Interaction

Catia Maria Dias Ferreira, Luciana Cardoso de Castro Salgado, and Clarisse Sieckenius de Souza

SERG / Departamento de Informática, PUC-Rio
Rua Marquês de São Vicente 225 – 22451-900 Rio de Janeiro, RJ - Brazil
{cferreira,lsalgado,clarisse}@inf.puc-rio.br

Abstract. This paper presents research carried out to explore the implications of giving users a specific vocabulary to express their perceptions and opinions about opportunities to make contact with cultural diversity in human-computer interaction. This two-step study is part of a broader research project that aims at investigating users' perceptions and reactions when interacting with cross-cultural systems. Our current findings point at the expressive power of the proposed vocabulary and the promising outcomes of using it in the interaction design cycle of cross-cultural systems.

Keywords: Cross-cultural evaluation, Culture, Cultural Viewpoint Metaphors.

1 Introduction

Today users can navigate almost anywhere in the Web (without national and cultural borders), making intentional or unintentional contact with foreign culture content. Thus the web has become a medium for intercultural encounters, a place where users have the opportunity to experience cultural diversity directly (when interacting with other users over social networks, for example) or indirectly (when interacting with web applications that carry cultural marks from a foreign origin).

Over nearly two decades of research with applications aiming to attend to the needs and expectations of people with different cultural and social backgrounds, we have worked to understand the impacts of culture on the users' experience and on user interface design. Studies have investigated, for instance, the adequacy of technologies [3], methodologies [4,8], theories and practices [3,5,6], as well as quality measures [10] created (and in some cases already validated) by and to a specific cultural group. There have also been new approaches to cultural sensitive interaction [7] and new theories of cultural influences in socio-technical systems [9].

Our research lies in exploring how cross-cultural systems, intentionally or not, express and promote *indirect* intercultural contact with signs and traces of foreign values, practices, heritage, and so on. The study presented here is part of a broader research about users' perceptions and reactions in such *indirect* intercultural encounters. At this stage, we concentrate on how users signify and talk about their experience using a well-defined set of conceptual metaphors [7] as part of their vocabulary.

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 314–322, 2013.

Cultural Viewpoint Metaphors (CVM) are a semiotic engineering [2] tool to support reasoning and decision-making about intercultural experience dimensions in HCI design. The general orientation we adopt is to view human-computer interaction as a special case of computer-mediated human communication where designers are telling the users, through systems interfaces, how, why, where, when and what for the system they have produced can be used. The users' interpretation and exploitation of the designers' *message* will define the kind of HCI experience that the system enables and ultimately whether the technology is successful or not.

In this paper we explore specific aspects of intercultural HCI experience, looking at the implications of introducing CVM in the users' vocabulary with which they can verbally express their perceptions and opinions when they faced opportunities for indirect intercultural contact. In the next section we describe our research. Then we discuss our current conclusions and present the main items in our future research agenda.

2 Research Aims and Method

We ran an empirical study with *Englishtown*[1] (ET), an online English school. This is an intercultural web application *par excellence* for all those who engage in learning English as a foreign language. We wanted to see: (i) if and how participants used CVM when verbalizing their intercultural experience with the website; and (ii) what they said they experienced. According to Salgado [7], the five conceptual metaphors lead us to think of interaction with cross-cultural systems as a journey, users being the travelers. There are five metaphors representing a *continuum* of cultural approximation which spans from the users' culture to a foreign culture (see Fig. 1).

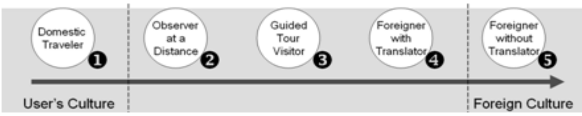


Fig. 1. Progressive cultural viewpoint metaphors [7]

At one end, the *domestic traveler metaphor* keeps users in cultural isolation, since they aren't exposed to material explicitly referenced to a foreign culture, and cultural mediation is therefore absent. With the *observer at a distance metaphor*, cultural markers from another culture are communicated as "bits of information". There are hints about the foreign culture for users that are interested in learning more about it. The *guided tour visitor metaphor* presents cultural markers from another culture as "illustration". Thus, aspects of foreign cultural features are exemplified and explained in the user's language, underlining contrasts between two cultures (the user's and the foreign one). The *foreigner with translator metaphor* allows users to experience the cultural practices from a foreign culture directly, with no other mediation than

[1] From now on we will refer to it as simply "the ET website" (`http://www.englishtown.com`)

linguistic translation of verbal content. Finally, the *foreigner without translator metaphor* treats the users as natives of the foreign culture, with no mediation whatsoever. We can thus see that cultural mediation is absent at the beginning and at the end of the *continuum* and that nature and degree of mediation varies in intermediate stages.

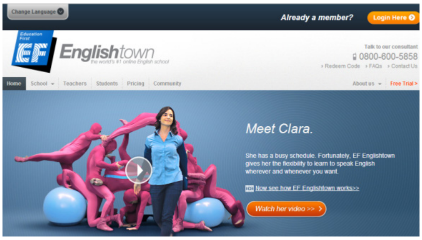


Fig. 2. ET website

To be sure, the ET website has not been designed with CVM, however, because it is clearly a cross-cultural application that exposes and exploits opportunities to communicate cultural diversity in the linguistic domain, we can use CVM to *talk about it*, as users or as HCI evaluators and designers. The ET mission, as clearly stated in the website "is to use technology to create a fundamentally better way of learning English. (…)As part of its endeavor to break down borders and remove barriers in language and culture, ET gives students more than just the ability to communicate in a new language" (see Fig. 2).

As shown in Fig. 3, our study was divided in two steps. In Step One, a Brazilian HCI evaluator (who knows CVM well) inspected the ET website using CVM to identify portions where different levels of cultural approximation were explored and interactive resources used to promote indirect intercultural encounters. Seven scenarios of inspection were then created for user sessions in the next step of the study.

In Step Two, five potential users were recruited to evaluate the selected portions of the ET website. None of them was an Englishtown student. Participants were further invited for a post-test interview. All of them were Brazilian, with college or university diplomas. They all had at least a basic knowledge of English and are interested in learning more about this language.

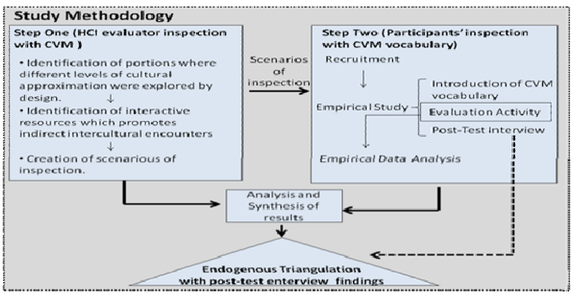


Fig. 3. Methodology

Empirical data collected in Step Two was influenced by the introduction of CVM vocabulary in the participants' expressive resources. Data analysis was carried out in three stages using discourse analysis techniques, a systematic exploration aiming to find out major meaning categories in discourse with intra-participant and inter-participant analysis. Firstly, we looked for evidence of each participant's usage and signification of CVM vocabulary during the evaluation activity, only. Secondly, results from the first phase were analyzed, contrasted and integrated with the findings from the inspection conducted in Step One. Finally, participants' interviews were analyzed separately, working as an internal triangulation required for validating our qualitative analysis. In this interview, we investigated how participants classified intercultural encounters according to the CVM framework and what perceptions they expressed with regard to such encounters.

The focus of inspection at Step One was on how ET explores English language teaching in different contexts: Business English; Social Protocol; Directions in English; and, Learning about the English language Places. The evaluator inspected each one guided by CVM, looking for interesting evidence of different degrees of cultural approximation. Selected portions were organized in accordance with the website's own categorization of interactive resources (articles, lessons, quizzes, videos) and their corresponding context. Figs. 4a and 4b show examples of selected material.

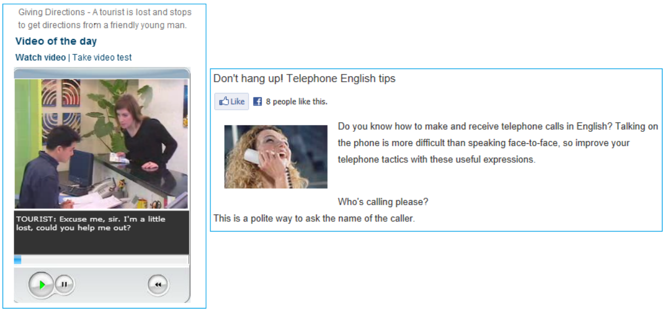


Fig. 4. (a) *Video* to learn **Directions in English** and (b) *Article* to learn **Social Protocols**

Next, the researcher analyzed the degree of cultural approximation promoted by selected interactive resources. We saw that the articles were being used to contrast two cultures with examples and explanations (guided tour visitor metaphor); videos typically sought to immerse students in the foreign culture (foreigner without translator metaphor); and quizzes and lessons led students directly into the foreign culture's context, providing them with linguistic support (foreigner with translator metaphor).

In Step Two, the five participants (P1, P2, P3, P4 and P5) were briefly introduced to the ET website and to the basic CVM concepts (with examples and illustrations). Then they were offered supportive material (for look up, during the activity) with a summary of the CVM vocabulary and the concepts associated with it. Next they listened to an explanation of the evaluation activity they should perform. There were seven scenarios for inspection, each involving one or more tasks. At the end, participants should describe and classify detected intercultural encounters enabled by ET, using CVM vocabulary they had just learned.

After completing the evaluation activity, during the post-test interview, participants reported verbally on what they had just done and experienced. This stage aimed at collecting evidence from the participants' discourse about the evaluation process with CVM vocabulary (*i.e.* how the proposed metaphors helped them, if at all, in talking about their perceptions and reactions; what they found easy or difficult to do; and what they learned with the activity).

3 Synthesis of Results and Triangulation

Results from the evaluation activity, in Step Two, pointed at two categories of meanings. In the first one, "Analysis of learning situations by using the CVM vocabulary", participants' discourse suggests different levels of usage and signification of the CVM vocabulary. We reached this conclusion based on two subcategories of meanings: (i) use of the basic CVM vocabulary to express perceived level of cultural approximation, without expanding or adapting it; (ii) free use of the CVM vocabulary to explain observed phenomenon spontaneously. For lack of space, we selected only some pieces of evidence to illustrate the kind of qualitative data we used.

In the first subcategory (i) P1, for instance, identified an intercultural contact as prescribed by the observer at a distance metaphor. P2, in turn, indicated another one as prescribed by the guided tour visitor metaphor:

P1: "[They are] *hints about the other culture* [communicated] *as information*".
P2: "*He* [the designer] *is contrasting cultures by means of examples*".

In the second sub-category (ii), we found evidence of participants' usage and signification of CVM vocabulary while explaining their own needs as users. See below that P5 also identified a case associated with the observer at a distance metaphor, but he expressed his perception in a more elaborate way:

P5:"[...] *he* [the designer] *is giving tips about what´s missing, simply saying what is right or wrong and pointing at the right answer. There isn´t an explanation, actually. I really expected some note. Why is it wrong? One does not understand why (…)".*

Evidence from P4 gives to us another great example of productive usage and signification, when he expresses how a learning situation with the English language places corresponds to the foreigner without translator metaphor.

P4: "*Not only because I do not have the translation or further linguistic support, but we are totally immersed through images, several attractive things, various sights being shown [...].I found it interesting because the narrator has an accent from* [the place]. *Then, everything they could do to show what it is like, how studying there should be like, parts of Sydney … they really immersed us there".*

The second category "Analysis of intercultural contact through language", presents some discourse excerpts with evidence that participants analyzed language as an important cultural component in the linguistic domain. The following subcategories of meanings show what they found: (i) politeness in language; (ii) translation used as linguistic support; (iii) language as a vehicle to expose foreign expressions; (iv) the close relationship between language and culture.

P4, for instance, in a context of Social Protocol learning situation, expressed his perception about subcategory (i).

P4: " [...] *the site explains that this is a very polite way* [to say something]. *If I just used* [the sentence] *the translator, I would not realize that this is a polite answer*".

The vocabulary of foreigner without translator metaphor (subcategory (ii)) was also used by P5 when faced with an opportunity to be completly immersed in another culture.

P5: "*[...] You are clearly inside their culture [...] They went heavy, just* [using] *slang*".

P3 pointed out that language may be used as a scaffold to give access to foreign material, as the foreigner with translator metaphor does (subcategory (iii)). P1, in turn, saw the challenges brought about by the close relationship between language and culture (subcategory (iv)).

P3: "*You are in another* [foreign] *culture, but you have a translation support*".

P1: "*It is a direct practicing of the foreign culture, without any linguistic support*". [It's an] *immersion in culture to request information. [...] I feel confused by the culture and language difference. The difficulty is because culture is related to language*".

In conclusion, from the subcategories of analysis presented above, we drew evidence that the methodology proposed here to access users' cultural perspectives worked. Participants were not limited to using, rigidly, the concepts of CVM. They literally "traveled" in metaphors' domain and went beyond the mere *tagging* of CVM's fixed expressions. They expanded, applied, adapted the concepts and expressed themselves freely with them. We thus conclude that the CVM vocabulary was visibly *incorporated* in participants' idioms.

In order to validate the qualitative research results presented above, we triangulated findings from the evaluation activity with results emerging from the post-test interview. After comparing and contrasting both results, we found consistent evidence that, just like in the evaluation study, participants again used and signified CVM vocabulary freely in their discourse. They also acknowledged the importance of taking cultural components into consideration when dealing with linguistic domains (as is the case with ET).

As evidence of the consistency among the results, we see in the following discourse excerpt that P2 clearly incorporated CVM terms into his own vocabulary, while talking about a learning situation:

P2: "*What caught my attention was the use of the Observer at a Distance as a honeypot: 'Look, here you will have something that will interest you. Do you know how to do this in another culture?'* ".

Furthermore, as was the case in the evaluation activity, during the post-test interview participants analyzed the challenges of using language as support. P4, for instance, explained that during translation many aspects of language are lost:

P4:"*If I do not have knowledge of the English language, I'll have access to this cultural information via translation and through somebody else's point of view. Thus, I'll lose a lot of the English culture, i.e., much of the language is lost during translation*".

Besides consistency reached in triangulation process, we also reached additional perspectives to our own research. Evidence suggests that participants found further opportunities to: (i) inspect a students' learning level using CVM's levels of cultural approximation; (ii) invite students to learn more about other culture; and, (iii) redesign the website. Regarding (i), the following excerpts talk about opportunities to monitor students' learning evolution:

P5: "*The beginner guy, he needs more information, he is open to receiving that* [new] *information. A person who is at a more advanced level, actually he requires less information and he also does not want to waste time with those observations* [he already knows it]*"*.

P3: "*You have a high degree of support while in your own culture and you go gradually removing it so that the person can get used to or interact more with another culture. The person is gradually supported with the transition, in order to be more immersed, moving across these areas* [the continuum with the five metaphors]*"*.

P2's excerpt, in turn, is about the power of the Observer at a Distance metaphor for inviting users to discover cultural diversity.

P2: *[…] "Do you know how to do this in another culture?"*

Finally, some participants saw opportunities to redesign the website.

P5: "*The quizzes could be focused on the "Guided Tour Visitor" metaphor because it does not matter just to know whether one is right or wrong, I am interested in knowing why this is an opportunity, they* [designers] *could improve the quizzes […]"*.

In conclusion, the triangulation showed the intensive use and signification of CVM vocabulary in the linguistic domain where the ET website is positioned. Participants produced very insightful reviews about their experience. Furthermore, participants helped the evaluator in seeing new elements in the HCI evaluation process, things that the evaluator herself had not found during the inspection.

4 Conclusion

This paper presented a qualitative empirical study where participants were invited to use the CVM vocabulary to express their perceptions regarding opportunities of intercultural encounters in a foreign language learning website. We wanted to see: if and how they used CVM when verbalizing their intercultural experience with the website; and what they said they experienced. The strategy of providing intentional vocabulary to be used in HCI evaluation is not new. The Cognitive Dimensions of Notations (CDN) framework, for example, has been proposed as "a **vocabulary** for design discussion" [1] and has been in use by a broad community of researchers. We were inspired by this approach that, according to the proponents, is meant to provide discussion tools "to help people who are not HCI experts in making quick but useful evaluations" [1]. The CVM vocabulary is comparable to the CDN one, as an HCI design and evaluation tool, except the Green and Blackwell take a cognitive orientation, whereas we take a semiotic one (brought about by the foundations of the CVM vocabulary[7]).

At this stage of our research, we have reached two main results. First, the CVM vocabulary was easily internalized and used by study participants. Second, participants produced very insightful evaluations of their experience with the website, including criticisms and suggestions. Together, these two findings mean that the CVM vocabulary generated empirical demonstration of its potential in the interaction design cycle of cross-cultural systems. CVM is a medium of expression and communication for users to qualify real or potential interaction experiences. CVM are, therefore, a promising support tool for participatory design practices.

All of our findings emerged from interaction with an application pertaining to the linguistic domain. Language, as we know, is a prime medium for cultural investigation, which means that the success of the reported research may be boosted by the specific domain where we worked. Thus, we now want to carry out a similar study in a non-linguistic domain. Likewise, and for sake of powerful comparison, we want to carry a similar study without offering CVM vocabulary for participants, in both linguistic and non-linguistic domains. This, we believe, will give us a very clear picture of the true power of CVM as a conceptual tool for intercultural HCI design and evaluation.

Acknowledgements. We thank the volunteer participants of our study. Clarisse de Souza and Luciana Salgado thank the National Council for Scientific and Technological Development (CNPq) and the Research Foundation of the State of Rio de Janeiro (FAPERJ) for financial support at different stages of this research project.

References

1. Blackwell, A., Green, T.: Notational systems: The cognitive dimensions of notations framework. In: Carroll, J.M. (ed.) HCI Models, Theories and Frameworks: Toward a Multidisciplinary Science, San Francisco, pp. 103–134 (2003)
2. de Souza, C.S.: The semiotic engineering of human-computer interaction. The MIT Press, Cambridge (2005)
3. de Souza, C.S., Laffon, R., Leitão, C.F.: Communicability in multicultural contexts: A study with the International Children's Digital Library. In: Forbrig, P., Paternò, F., Pejtersen, A.M. (eds.) Human-Computer Interaction Symposium. IFIP, vol. 272, pp. 129–142. Springer, Boston (2008)
4. del Gado, E.M., Nielsen, J. (eds.): International Users Interface. John Wiley & Sons (1996)

5. Irani, L., Vertesi, J., Dourish, P., Philip, K., Grinter, R.: Postcolonial Computing: A Lens on Design and Development. In: Proc. ACM Conf. Human Factors in Computing Systems, CHI 2010, Atlanta, GA, pp. 1311–1320 (2010)
6. Maunder, A., Marsden, G., Gruijters, D., Blake, E.: Designing Interactive Systems for the Developing World – Reflections on User-Centred Design (2007)
7. Salgado, L.C.C., Leitão, C.F., de Souza, C.S.: A Journey Through Cultures: Metaphors for Guiding the Design of Cross-Cultural Interactive Systems. Springer Publishing Company, Incorporated (2012)
8. Vatrapu, R., Pérez-Quiñones, M.: Culture and Usability Evaluation: The Effects of Culture in Structured Interviews. Journal of Usability Studies 1(4), 156–170 (2006)
9. Vatrapu, R.: Explaining Culture: An Outline of a Theory of Socio-Technical Interactions. In: Proceedings of the 3rd ACM International Conference on Intercultural Collaboration, Copenhagen, Denmark (2010)
10. Winschiers, H., Fendler, J.: Assumptions considered harmful: The need to redefine usability. In: Aykin, N. (ed.) Usability and Internationalization, Part I, HCII 2007. LNCS, vol. 4559, pp. 452–461. Springer, Heidelberg (2007)

Characteristics of Elderly User Behavior on Mobile Multi-touch Devices

Susumu Harada, Daisuke Sato, Hironobu Takagi, and Chieko Asakawa

IBM Research - Tokyo
NBF Toyosu Canal Front Building, 5-6-52 Toyosu, Koto-ku, Tokyo 135-8511 Japan
{haradas,dsato,takagih,chie}@jp.ibm.com

Abstract. Smartphones and tablet devices have been rapidly proliferating, and multi-touch interaction, powerful processors and rich array of sensors make these devices an attractive service platform for older users. While there is an increasing number of work investigating the issues that elderly users experience through their interaction with mobile devices, most have focused either on evaluation of low-level interaction characteristics or on qualitative survey. Therefore, we conducted a user study with 21 elderly participants to analyze the needs and issues faced by this user group under naturalistic usage scenarios. Specifically, we interviewed each participant about their experiences, had them perform various practical tasks using our custom testing application, and analyzed the operation logs using our custom visualizations. Based on our results, we summarize the types of issues observed, present design considerations for the applications studied, and future research directions.

Keywords: Mobile, Multi-touch, Smartphones, Tablet, Aging, Elderly.

1 Introduction

Over the past several years, mobile multi-touch devices such as smartphones and tablets have been rapidly proliferating. While the primary growth has been among younger users, there is a growing number of elderly people who are beginning to adopt such new technology [1]. The direct manipulation interaction afforded by their multi-touch displays, their ability to run numerous applications both on the device as well as over the network, combined with powerful processors and a rich array of sensors, make these devices an attractive platform for making a wide range of services available to benefit older users.

While the multi-touch interfaces enable intuitive direct-manipulation interaction that mimic real-world metaphors, they also introduce a number of potential challenges, such as non-intuitive multi-finger gestures, unexpected sensitivity of the touch surface, and a conceptual model that differ significantly from traditional desktop computers and other preceding technologies.

Younger users may be able to quickly learn to navigate around such challenges through trial-and-error and relying on their mental models of recent technologies, but the hurdle may be bigger for older users. A study by Docampo Rama, Ridder, and

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 323–341, 2013.

Bouma [2] indicates that generational difference in exposure levels to various technology during one's formative years (before the age of 25) has a significant effect on the performance on and the level of understanding of new technologies. A survey with 131 respondents from three age groups (20-49, 50-64, 65+) conducted by Leung and others [3] suggests that significantly fewer older people choose trial-and-error as a method for learning new technology compared to younger people, opting for more traditional methods such as instruction manuals despite citing difficulties using them. It is clear that the needs and characteristics of the older population are quite different from those of the younger population in the context of mobile device usage. In our investigation, we focus on the issues surrounding the usability of such devices from the perspective of elderly users.

There have been many works seeking to better capture the needs and characteristics of older users and their interaction with mobile devices. On the one hand, there have been various interview and questionnaire studies [3–7] that have identified characteristics unique to the elderly user population. While these interview and questionnaire studies provide valuable insight into elderly user's attitudes towards and perceptions of mobile devices, they do not reveal much about the interaction tendencies of and challenges faced by elderly users throughout their actual use of such devices.

On the other hand, there have also been works that sought to capture more concretely the patterns and tendencies of elderly users' interaction with mobile touch-screen devices through measurement and analysis of low-level interaction metrics such as task speed and accuracy [8–13]. Most of these works involved experiments in which the participants performed primitive actions such as tap or swipe gestures in isolation within their custom test applications. While these studies provide concrete data, the data may not be representative of what users experience in actual usage of mobile devices, since real tasks require not only primitive gesture skills but also skills to develop strategy to complete tasks by effectively selecting a series of gestures in real-time, in-situ, within the context of full application user interfaces.

In our work, we build upon the above studies by combining an observational study approach with an experimental approach that include detailed instrumentation of elderly users' interaction with mobile touch-screen devices as they perform tasks within realistic application contexts. Our aim was to uncover issues that elderly users may encounter during their typical usage of smartphones and tablets, such as errors, unexpected results, standstills, ineffective operational strategy and so on, and delve deeper into each issue by analyzing the instrumented log data.

We conducted a user study with 21 elderly participants with diverse experience levels, from first time users to active intermediate level users, to observe and analyze specific issues and challenges that they experience when using mobile touch-screen devices. We built a custom experiment application for multi-touch smartphones and tablets that mimicked the appearance and functionality of actual applications, augmenting them with instrumentation capability to capture and log all multi-touch events and application state changes. Our focus, however, was less on statistically analyzing aggregate results but more on carefully observing interaction characteristics such as errors, unexpected results, operational strategies and so on, of each individual participant. We also conducted interviews with each participant to solicit their comments and better understand their perceptions and past experiences with mobile devices. Based on these results, we present a number of design considerations and future research directions for creating smartphone and tablet devices and applications that are more accessible and easily adoptable by elderly users.

Our contributions are as follows:

- Quantitative analysis of elderly users' interaction behaviors within realistic application settings
- Qualitative analysis of elderly users' specific tendencies and obstacles encountered during their interaction with multi-touch smartphones and tablets

2 Related Work

There are many prior work that have investigated elderly users' perceptions and use of traditional mobile phones [14–17], as well as those that have examined elderly users' interactions with large touch-screen displays, but not much work has yet looked into elderly people's use of mobile touch-screen devices such as recent multi-touch smartphones and tablets.

2.1 Elderly Users and Mobile Phones

A number of interview and questionnaire studies have been conducted to investigate the general perceptions and practices of elderly users regarding mobile phones, and several resources provide a comprehensive overview of the literature [1, 18]. We present some of the representative works here.

Kurniawan [4] conducted a set of expert interviews with two domain experts, focus group discussions with 14 elderly participants, and an online questionnaire that resulted in 100 responses from people over 60 years old. The study reports on various findings about elderly people's perceptions of and issues with mobile phones, such as their preference for features that aid their declining functional abilities. Leung and others [3] also conducted a large-scale online survey with 131 respondents across three age groups (20-49, 50-64, 65+) to identify unique tendencies among the older population, including their preference for traditional methods such as instruction manuals for learning new technologies over trial-and-error. Van Dyk, Renaund, and van Biljon [7] conducted an extensive interview with 147 mobile phone users between the ages of 60 and 89 to construct a prioritized checklist of features that should be considered in the design of mobile phones catered to the elderly users.

There have also been some observational studies as well as experimental studies that delve into more detail about the elderly's use of such technologies. Renaud and Biljon [6] used worth-centered design process to interview and conduct a participatory design experiment with elderly users, resulting in a prototype mobile phone design. Weilenmann [19] conducted an observational study of elderly users learning to enter text on mobile phones, highlighting specific issues they encountered. Ziefle and Bay

[13] conducted comparative evaluation of two mobile phones of different complexity between younger and older participant groups. Jastrzembski and Charness [8] proposed and validated the parameters of a Model Human Processor for older adults.

While these studies provide valuable insight into the overall tendencies and preferences of elderly people as well as their usage characteristics within the context of mobile phones in general, they have not focused on issues that surround the more recent mobile multi-touch devices.

2.2 Elderly Users and Touch-Screen Displays

Touch interaction is one of the key features that differentiate recent smartphones and tablets from prior mobile devices. Leonardi and others [20] conducted experiments with a tabletop touch panel interface and found that the direct interaction metaphor was easy to understand and had a pleasing effect that attracted and motivated elderly participants in their study. Lepicard and Vigouroux [21] had a group of younger and older participants interact with their test application on a 12” Tablet PC and found that two-handed touchscreen input was difficult for elderly users. There have also been other observational studies of elderly people’s use of touchscreen interfaces [22, 23], and a survey of other work is provided by Caprani, O’Connor, and Gurrin [14].

While the results from these studies offer insight into the issues that elderly users experience with touch-screen devices, there are several key differences between tabletop and desktop touchscreen devices and mobile multi-touch devices that warrant further investigation looking specifically at the latter category of devices. For one, the smaller form factor of mobile devices mean that the users are often holding the device with one hand while they touch the screen with the other hand or even the same hand. This factor, combined with the multi-touch sensing, can potentially lead to greater chances of inadvertent or accidental touches of the screen by the holding hand or some other part of their manipulating hand. Furthermore, the smaller screen estate can lead to the need for more touch manipulation to view the desired information as well as potential errors when touch targets are too close or too small. Therefore, more research is needed to investigate the more recent multi-touch smartphones and tablets.

2.3 Elderly Users and Mobile Multi-touch Smartphones and Tablets

While still few in number, there have been recent works that have specifically focused on mobile multi-touch smartphones and tablets as the target of their study. One part of the extensive study conducted by Leung and others [3] involved a field study observing two middle-aged and four older users learning to use smartphones. Kobayashi and others [9] observed 20 participants in their 60s and 70s using various applications on multi-touch smartphones and tablets. They also conducted controlled experiments involving execution of primitive gestures on smartphones and tablets, finding that the elderly participants’ performance of dragging and pinching gestures improved after one week of practice but not tapping. Leitão and Silva [10] used a custom “game” on a multi-touch smartphone to analyze the effects of target size and spacing on tapping and swiping tasks. Nicolau and Jorge [11] evaluated typing performance and typing patterns of 15 participants over the age of 65 using a virtual keyboard on a multi-touch smartphone and tablet.

Our goal was to build upon these works by observing and looking for specific challenges, stumbling points, and possible coping strategies employed by individual users. While our results do not provide statistical significance or sweeping generalizations, we believe that the concrete observations, both qualitative as well as quantitative, situated within the context of realistic application usage, provide valuable insight into specific issues that elderly users may encounter in their typical interaction with multi-touch smartphones and tablets.

3 User Study

We conducted seven half-day-long user study sessions with a total of 21 elderly participants to investigate their perceptions of smartphones and tablets as well as specific usage characteristics as they performed tasks on the devices. Our goals were:

1) to better understand elderly users' perception of smartphones and tablets,
2) to identify "realistic issues" encountered by elderly users as they interacted with real and "realistic" apps,
3) to quantitatively analyze the issues identified in 2) to see if we could systematically identify and potentially circumvent such issues, and
4) to uncover design considerations that could inform design of more senior-friendly smartphones and tablets.

3.1 Participants

We recruited participants for our study through email announcements distributed to retiree mailing list for a large corporation, as well as to local computer classes for seniors. The condition for participation was specified as being 60 years of age or older with some interest in smartphones and tablet devices.

A total of 21 participants ranging in age from 63 to 79 (12 females and 9 males) took part in our study. Fifteen of the participants were members of the local computer class, and six were from the retiree group. The participants were divided into seven groups, each consisting of two to four members, so that we could conduct a focus-group-style discussion. Each group was invited to our lab for a half-day session, and each participant was compensated for their time with a gift valued at approximately $25. All of the participants owned a mobile phone. Eight of them owned multi-touch smartphones while 14 owned a tablet, and six owned both devices.

3.2 Procedure

Each of the half-day user study sessions was structured into two parts: 1) open discussion about each participant's experiences with and perceptions of smartphones and tablets, and 2) execution of controlled tasks on smartphones and tablets. The sessions were audio recorded for later analysis, and with participants' permission, video was also recorded to capture the user's hands operating the devices during various tasks. During each session, two experimenters were present, one to primarily engage with the participants, and another to take notes and to tend to the video cameras.

Open Discussion. At the start of the user study session, we first had all participants individually fill out a questionnaire soliciting demographic information as well as specifics about their experiences with smartphones and tablets. In particular, we prepared different questionnaires for those who did or did not own a smartphone or a tablet. To those who did not own the devices, we asked whether they had considered owning one, and if so, what factors had interested them and what factors had prevented them from actually owning one. To those who owned the devices, we asked for their reasons for purchasing them and what they most use them for. After the questionnaires were filled out, about 30 minutes was spent on open group discussion about each participant's experiences and perceptions about smartphones and tablets as well as their current mobile phones. The discussion was led by the experimenter based on a prepared set of guiding questions, but the participants were free to expound upon or add any stories about their experiences. The group discussion format was adopted so that the participants were able to engage in a more casual conversation, feeding off of each other's responses and stories.

Controlled Task Experiment. In the controlled task experiment session, we had the participants perform various tasks on both the smartphone and tablet using the data collection application we built to gather logs of quantitative usage data. The purpose of this session was to attempt to gain an objective view of any tendencies or unique characteristics in the interaction styles of each participant in using various touch-operated apps. The primary objective of this investigation was not to obtain statistically significant aggregate data, but rather to observe and analyze the usage patterns of individual participants so that we may gain specific insight into concrete manifestations of errors, operation difficulties, and other issues. Ideally we would have liked to evaluate the usage of real applications under actual usage scenarios, but we resorted to an in-lab study to more closely observe the interaction issues as they happened.

Each participant performed in an order of Address Book, Phone, and Map tasks on both smartphone and tablet. The order of devices was counterbalanced. All participants performed the same set of tasks on both devices. For the Phone task, they input three 11-digit phone numbers printed on a sheet paper. For the Address Book task, the participant was presented with a same list of 160 names but with different initial starting position in the list, and were prompted to search for and tap on a particular name. A total of 10 pre-chosen names were presented one by one and were shown at the top of the application. For the Map task, the participants were asked to find three locations (Hokkaido, Okinawa, and England) on a map, zooming in to fill up the screen with the target island as the completion condition for each trial. The map was reset to show the Tokyo Bay area at the beginning of each trial. While not all apps required multi-touch gestures, we wanted to see if the multi-touch-enabled screen would lead to various errors caused by unintentional touches during the tasks, a common symptom observed especially among novice multi-touch device users.

3.3 Apparatus for the Controlled Task Experiment

For the controlled task experiment, we used Apple's iPod Touch (fourth generation) and iPad (third generation). Both devices were running iOS version 5. We built a custom test application that included three application interfaces that closely mimicked the appearance and behavior actual applications found on standard iOS devices (Phone, Address Book, and Map). The test application logged all multi-touch events and application events for later analysis.

Phone Address Book Map

Fig. 1. Screenshots of the three application interfaces in our test application

Phone Application Interface. The phone application interface consisted of the standard telephone keypad, and an entry display at the top showing the numbers entered. The actual visual of the user interface was directly copied from the original application. The app was set up such that after the participant entered the correct phone number for a particular trial and tapped the dial button, the screen returned to the main screen. If the user tapped the dial button but the phone number entered was incorrect, a dialog was displayed to prompt the user to re-enter the correct number.

Address Book Application Interface. The address book application interface consisted of a vertical list of contact names sorted and grouped alphabetically, with an index bar on the right side of the screen that support jumping to the section of the list corresponding to the tapped index letter. During each task trial, the participant was first presented with a dialog displaying the target name to find in the contact list. After dismissing the dialog, the target name remained visible at the top of the screen as the user scrolled through the list.

Map Application Interface. The map application interface simply presented the standard iOS map control, with the ability to use two-finger pinch-in and pinch-out gestures to zoom out and in, respectively, as well as to pan the map using either one or two fingers. Double tapping with a single finger to zoom in by a fixed amount and single tapping with two fingers to zoom out by a fixed amount were also possible.

4 Controlled Experiment Results

4.1 Phone Application Task Results

Position of Touches. Participants completed each trial on average in 13.2 seconds on the smartphone and in 11.9 seconds on the tablet. For each of the 63 (3 x 21) trials on smartphone and tablet, the numbers of incorrect touches were 32 and 15 in total, respectively. Figure 2 shows the plots of correct touches and incorrect touches made by all participants on both devices. With the tablet, five participants made incorrect taps with their palm, which can be seen at the bottom-right of the plot for the tablet. With the smartphone, three participants made an incorrect "call" while intending to tap "0". Most of the other incorrect taps were due to misreading the phone numbers.

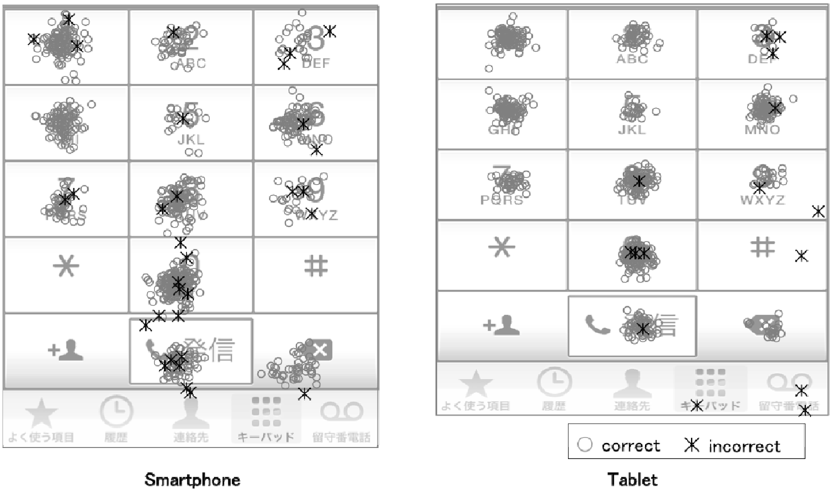


Fig. 2. Correct and incorrect touches in the Phone task on smartphone and tablet

We observed P13 input one extra digit with an unexpected repeat of the "8" button. By checking the recorded video for the trial, she did only tap once on the "8" button but the system accidentally recognized two distinct taps within 0.13 seconds. Also many unregistered taps were observed, in which participants physically tapped on the screen but the system failed to recognize them.

Not Confirming before Moving on. There was an instance when P13 tapped the "call" button with one too many digit entered, and the dialog prompted her to correct the entry. She then tried to re-input the number from the beginning but only tapped the delete button once, not checking the number display at the top, resulting in an even longer sequence of numbers. P04 and P15 also got the dialog prompt during the task by tapping "call" button with too few digits of input. After they dismissed the dialog, they also tried re-entering the digits from the beginning without deleting or checking the display at the top of the screen.

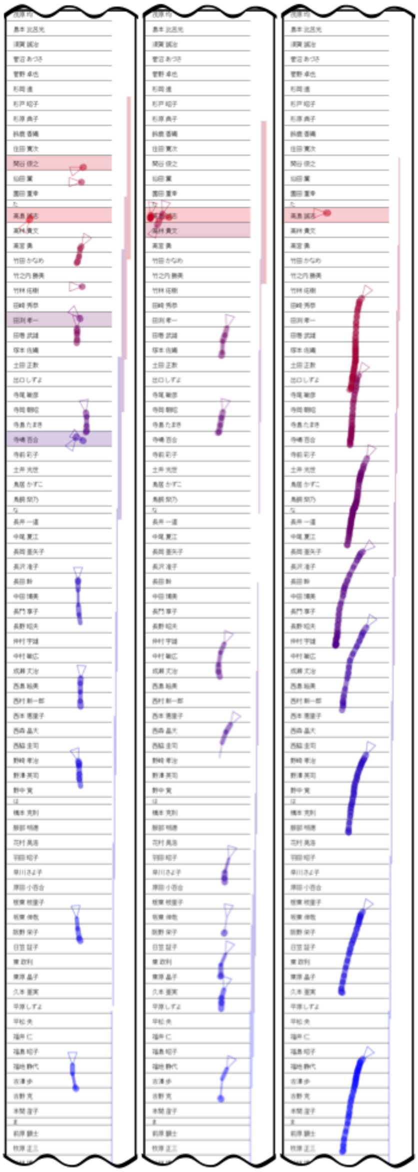


Fig. 3. Trajectories of all gestures within representative trials for the scroll task on the smartphone across three participants. (Nudge-flicking: P11 and P14, Stroke-flicking: P18)

4.2 Address Book Application Task Results

Participants completed each trial on average in 9.7 seconds on the smartphone and in 8.9 seconds on the tablet. P09 and P11 made 15 and 12 miss-taps, respectively, significantly more than the maximum of four miss-taps made by the other participants. Most of the mistakes by P09 on the tablet were miss-selection of list item while intending to tap on the index bar. All miss-taps by P11 were on the smartphone, and seemed to be due to her hand tremor.

Difficulties with the Index Bar. P4, P5, P9, P18 and P23 experienced a number of instanced in which they unintentionally manipulated the index bar. In particular, the participants were especially confused when they were attempting to drag the list vertically but instead accidentally started their gesture on the index bar. As they moved their finger, the list jumped discretely across the index groups instead of smoothly following the participant's finger movement. P4, P5 and P9 experienced these issues even though they were aware of the index bar function previously. P4 also exhibited confusion when she tapped the index bar my mistake when attempting to tap the target row.

Similar to the case with the Phone application interface, we also observed a number of unintentional touches primarily by the participants' palms. Due to the index bar being on the right side of the screen and all participants being right handed, this often led to the index bar being accidentally tapped, resulting in the list jumping to a grouping unexpectedly. As a result, P23 commented, "the screen just 'flew away'… I just touched it lightly just like this, and then it just flew away somewhere…"

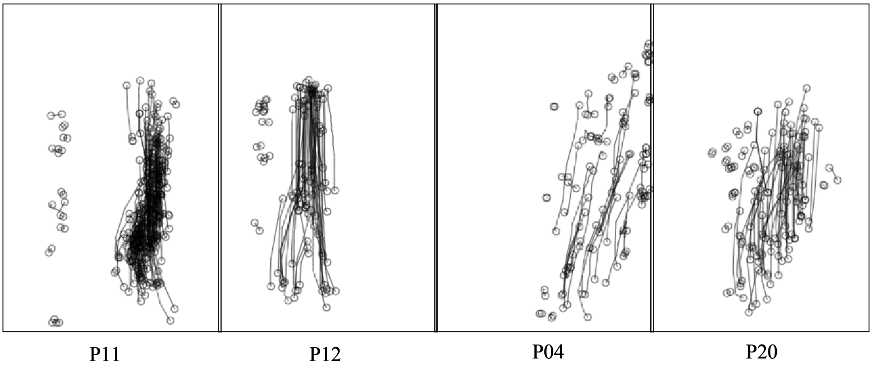


Fig. 4. Trajectories of all gestures for the Scroll task on the smartphone

Nudge-Flicking Instead of Stroke-Flicking. There were distinct differences among the participants in their finger movements for scrolling the list up or down. It is typical for experienced users to use short, quick "flick" gestures to "throw" the list towards a given direction to take advantage of the inertial scrolling to scroll over a long distance, and to use controlled "dragging" gesture to keep the list pinned below the firmly touching finger to move the list by a small amount. Figure 3 shows a visualization of users' gestures during a particular trial of the Address Book task. Each stroke represents the user's gesture on the screen. Triangles at the beginning of each stroke represent the stroke direction, and rows that were tapped, including both correct and incorrect selections, are shown filled. While most users where able to utilize inertial scrolling, P11 and P14 in particular used extremely short strokes for flicking, only moving the tip of the finger to perform the gesture, in contrast to participants such as P18 who exhibited a much longer flicking stroke, moving the entire hand or even the arm to perform the gesture.

Figure 4 shows the trajectories of all gestures across all trials on the smartphone by P11, P12, P04, and P20. P11 and P12 taped on the left side of the smartphone when selecting a name on the list. Each list item has a full screen width, and all participants except P4 and P20 tended to tap on the side of the screen with the names displayed.

4.3 Map Application Task Results

Participants completed each trial on average in 23.8 seconds with the smartphone and in 23.7 seconds with the tablet. Figure 5 shows a visualization of users' operations as they navigated from Tokyo to England on the map (P6 on the tablet and P15 on the smartphone). Each rectangle in the visualizations shows what the device showed when they performed the corresponding gestures and the solid lines within the rectangles represent the gesture trajectories. They used zoom-out gestures but the zoom-out level was not enough to explorer the world map, and they tried to find the goal (England) by executing numerous swiping gestures.

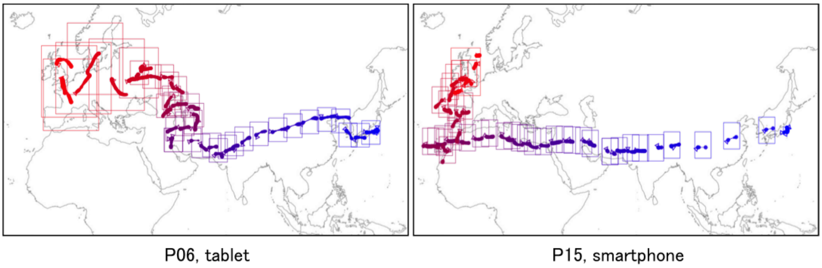


Fig. 5. Operation sequence for a particular Map task trial by P06 and P15

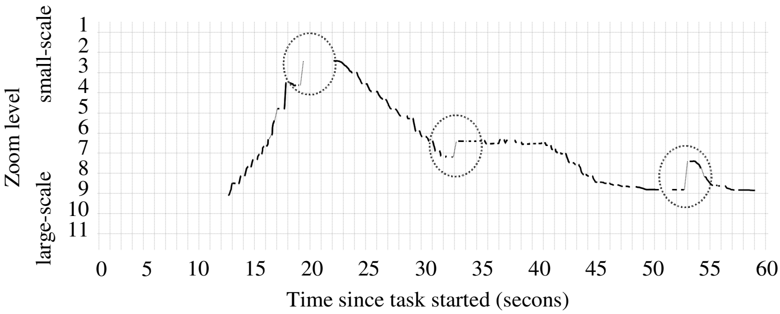


Fig. 6. Zoom level sequence for a particular Map task trial by P19

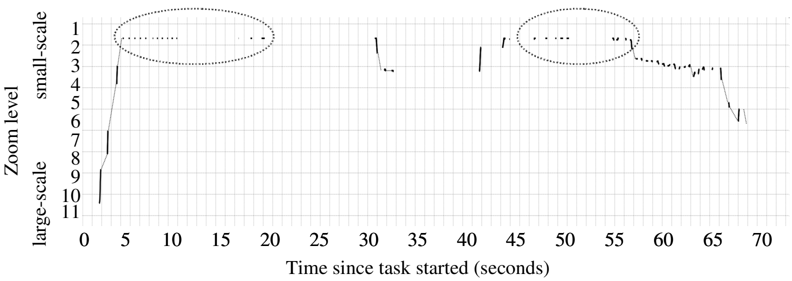


Fig. 7. Zoom level sequence for a particular Map task trial by P17

Figure 6 and Figure 7 visualizes the zoom level sequence of the task by P19 and P17, respectively. Vertical axis shows the zoom level from world (small-scale) to town (large-scale) and horizontal axis shows the time since the task started in seconds. Thick lines indicate zoom level change while touching and thin lines indicate zoom level change without touching (inertial, double tap, or single tap with two fingers). P19 used two fingers for both zooming and panning and each gesture was very short, causing unexpected zoom-outs during operations. Three dashed circles in

Figure 6 indicate where the map recognized single tap with two fingers to zoom out for a certain step. This type of miss operations was also found in P13, P14, P15, P17, P20, P21 and P23. Two dashed circles in Figure 7 show that P17 attempted many pinch-in (zoom out) gestures at the minimum scale. He was aware that something was not going as expected (i.e., the map was not zooming out any further than its current level), but he did not stop executing the zoom-out gesture.

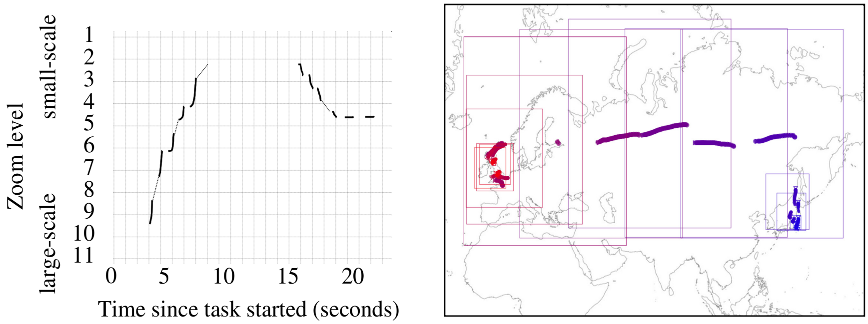


Fig. 8. Operation sequence for a particular Map task trial by P23

Figure 8 shows a good example of operation by P23 from Tokyo to England. He could zoom out smoothly and easily found the target after a few swipes and then zoomed into the target. By comparing such participants' log traces, it becomes clear that there are distinct differences in the characteristics of the performed gestures as well as their strategy in performing the task of locating a target on a map.

The smaller and incremental pan and zoom gestures were observed more among those who had less experience with smartphones and tablets (i.e., those who did not own the corresponding devices), so we looked into the differences in the ratio of zoom operations of varying magnitudes, grouped by device experience (Figure 9a). A similar analysis was performed for panning movement distance as well (Figure 9b). Figure 9a shows that people who do not own smart devices tend to use smaller scale factor of zoom-in and zoom-out. Figure 9b shows that people who do not own smart devices tend to use shorter panning, most of the panning were less than 300 pixels, while people having smart devices use various length of panning.

Comparing to the other two apps, many difficulties of operations were observed with map application. P6 and P19 use more than two fingers for zoom-in and zoom-out. Unfortunately, since the tablet has a function to close application by four fingers pinch-in gesture, the experiment app was terminated sometimes by the four fingers gesture when they wanted to zoom-out on the map. Two hands were used to perform for pinch-in/out gestures by P17 and P20 on smartphone and tablet respectively. P17 had difficulty zooming with two fingers on smartphone. Many participants performed a single tap gesture on the map instead of a double tap gesture when they saw their target location at a low zoom level, seemingly in an attempt to zoom in to the target location.

5 Interview Results and Observations

Here we report on other observations across the experimental tasks and through our conversations with the participants.

5.1 Unexpected Touch Screen Responses

There were a number of instances when a participant either intended to touch the screen but the system did not register (unregistered touches), or did not intend to touch the screen but the system registered a touch event (unintentional touches). Such unexpected responses seemed to be a major cause for frustration and confusion expressed by the participants.

Regarding unregistered touches, the primary cause seemed to be that the participants' fingers were too dry for the touch-sensitive display to detect. P4, P5, and P6 mentioned frequently having problems during their daily life with dry finger resulting in ATMs and other touch screens not responding. P14, P18, and P20 all commented during their experimental tasks that their fingers felt dry and that the device seemed not to be responding to their touches.

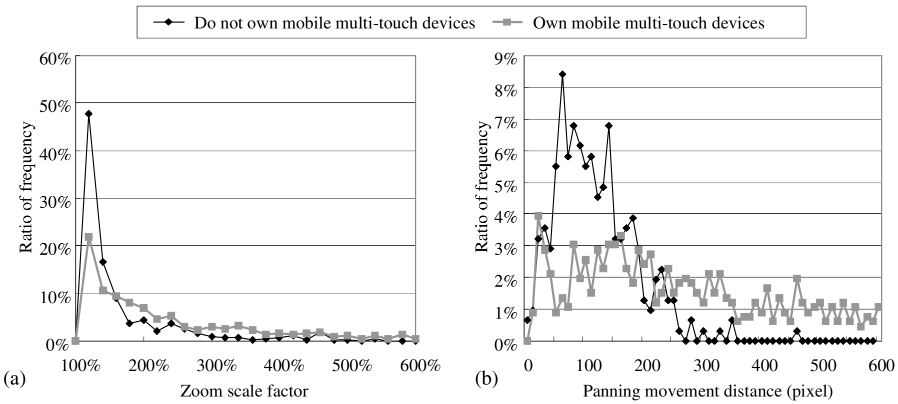


Fig. 9. (a) Comparison of frequency of zoom scale factor in the Map task across users who do and do not own mobile multi-touch devices. (b) Comparison of ratio of panning gesture lengths in the Map task across users who do and do not own mobile multi-touch devices.

Unintentional touches were also observed across many participants. There were several patterns of unintentional touches. First was when the participant's non-controlling hand that was holding or resting near the device accidentally touched the screen surface. Another was when the participant was hovering above the touch screen surface contemplating what to tap, and got too close to the surface of the display and triggered a touch event. Finally, there were cases when the user was performing some gesture such as a single tap or a swipe, during which some other part of their hand also accidentally contacted the display surface, resulting in multi-touch events being registered.

P10, P13, and P14 mentioned that they often experience situations in which unintentional touches are triggered by the sleeve of their clothing making contact with the touch screen at ATMs or when using their tablet. While we do not know for certain whether it is actually the sleeve of their clothing that was making contact with the touch screen or a part of their hand, but it is clear that unintentional touches are quite a common phenomenon that is noticed by these elderly users. P6, P11, P12 and P22 all mentioned that they experienced difficulty with such unintentional touches when they first purchased their touch screen devices but that they had now gotten used to it.

5.2 Not Seeing the State of the Entire Screen

A number of participants tried to start scrolling before dismissing the popup dialog in the address book task. In the phone app, when using the backspace button, some users seemed to assume that it deleted not just the most recent digit but all digits entered so far. When such a user miss-entered a digit in the middle of a sequence, they tapped the delete button once but then proceeded to reenter the entire number sequence from the first digit, and not being aware of their mistake as they did not seem to be checking the number display at the top.

During the interview session, we observed some users entering text into a search textbox. During this process, these users were focused on the on-screen keyboard keys, and were not looking at the actual text being entered in the text box, which was at the upper left corner of the screen. Unlike physical keyboard where touch-typing is possible due to the tactile feedback from the physical keys, touch-screen keyboard inherently requires user to direct visual attention to each key the user is about to tap. For elderly, seems especially challenging to shift attention back and forth from keys to the textbox where text is being entered. As a result, they type away until they finish their phrase, and only then go to check what they typed, which may include surprising results due to mistyped characters and auto-suggestions.

5.3 “Unfriendly” Interface

During the interview session, a number of participants tried entering text into a memo application using the on-screen keyboard. There were several keyboard layouts available, and to switch among them, a particular button had to be pressed. However, if the user stays in contact on that button for slightly longer than a quick tap, to bring up a popup menu with a list of available keyboards is shown. The popup, however, automatically fades away after a few seconds. P3, P4 and P5 had difficulty with this, as they would tap the button once, and before they could finish reading the list of options and tap on a selection, the popup faded away.

Tap-and-hold menus are tricky, especially because by holding the finger down, it may be visually covering the available options. Tap should show the menu and keep it shown, but the issue is, how to dismiss the menu without making a selection. Tapping “outside” the menu should be used, but sometimes, since the background behind the popup menu is still highly visible, one may be reluctant to tap on the “background” for fear of activating something else. A clearer indication of which part of the interface is actionable at any time may be needed.

5.4 The "Aha" Moments When Being Shown a New Technique

After the controlled task experiments were completed, the experimenters debriefed the participants by describing and demonstrating some techniques for performing some of the interactions more accurately and precisely. There were a number of instances when users expressed particular surprise and joy at learning about such new method of interaction that was more effective than what they had been used to. For instance, P4 and P5 were excited to be shown that the device can also respond with the side of the finger instead of just the tip. P10, P11, and P12 also expressed excitement at learning how to perform the pinch-in gesture on the map by not flicking the two fingers together but by assuming that the map is like a rubber sheet in which the points under the fingers remain fixed as the fingers are moved.

As elderly people may be less reluctant to experiment [3], it may be good to show various alternative interaction approaches, preferably gradually and repeatedly, to accommodate for their slower learning pace.

6 Discussion

This section describes a number of design considerations for enhancing the three applications we used in our study, based on the above results, to address some of the issues encountered by our elderly participants.

6.1 Phone Application Interface

As described in Section 4.1, the main issues experienced by the elderly participants in the phone application interface were accidental taps of buttons adjacent to the target buttons, especially problematic between the "zero" and the "call" buttons, and also not being aware of the top part of the screen showing the phone number. The issues of unexpected touch screen responses are one of major issues across applications (Section 5.1).

One way to mitigate these issues may be to provide better feedback through both audio and visual. The phone application currently provides dial-tone feedback whenever a key is pressed, but this could be enhanced by having the tapped digit be spoken out. Work by Sato and others [24] have shown that such spoken feedback during online web form entry helped elderly users feel more confident and reassured. Such feedback mechanism will help various situations such as misrecognition caused by "dry" finger or cloths (Section 5.1) In addition to sound feedback, visual feedbacks may be effective. The phone application also highlights the pressed button. In addition to this visual feedback, showing a fading "fingerprint" on the display to show the actual touch position may be helpful for the users. This could be combined with auditory icon and/or color feedback to indicate number of touches detected.

Furthermore, the application could determine whether the input number is a valid phone number or not to change behavior of the "call" button dynamically. For example, if the phone number is not valid, the system can show a confirmation dialog to the user. More implicit way to handle the situation may be for the system to ignore the press of the "call" button once but letting the second press invoke the actual call.

6.2 Address Book Interface

Although using the index bar to navigate to an entry in a list can be effective (Section 4.2), a number of our participants experienced difficulties in manipulating the index bar. Some participants unintentionally tapped on a row entry instead of on the index bar, some participants operated index bar before dismissing the popup dialog (Section 5.2) and some participants touched the index bar at the beginning of their swipe gesture when scroll the list (Section 4.2). There may be several reasons for these mistakes, including the index bar being too narrow, the touchable area of the index bar not being visible (it temporarily becomes visible while the user is touching the index bar), and the index bar being laid out on the right edge of the screen. One possible solution for the current index bar design is to place the bar on the left edge of the screen for right-handed users and making the labels bolder and more visible. If this design is not acceptable from an aesthetic point of view, the index bar may be better off being removed altogether. As the scroll list component handles both tap gestures as well as swipe gestures, there is a greater possibility that the user's gesture may be misinterpreted compared to the phone application that only handles tap gestures.

6.3 Map Interface

Under the map interface, there are more gestures available than with the address book interface, including pinch to zoom, double tap, two-finger single-tap, etc. While our map interface did not implement them, there are other map applications that support even more gestures such as tap and hold for dropping pins on the map and using two fingers to rotate the map, making maps one of the more comprehensive apps in terms of the types of gestures supported. Furthermore, the notion of an "incorrect" operation on map interface is not as clear as with the phone or the address book interfaces.

As a possibility for enhancing the map application interface for elderly users, it may be desirable to support the ability to disable or enable various gestures to match the skill level of the user. Furthermore, a training mode that teaches the user about each available gesture and assesses users' skill of gestures by analyzing their panning and zooming behaviors (Figure 9) allows the user to practice them and to be aware of the correct concept of map operation (Section 5.4). It would not only benefit the map application interface but other applications in general as well.

We also observed a participant repeatedly attempt to pinch out beyond the maximum supported zoom level (Section 4.3), but such a behavior of repeatedly attempting ineffective gestures was not seen in the address book interface. The reason could be that in the address book interface, when the end of the list is reached, any attempt to scroll further results in a bouncing animation where the list bounces back to its limit position.

By adopting a similar animation feedback for the map's zoom function, it could more clearly communicate to the user that the attempted action is ineffective.

7 Conclusion

We investigated the issues and challenges that elderly people may encounter when using multi-touch smartphones and tablets by conducting a user study with 21 elderly participants. The user study sessions comprised of open discussion about each participant's experiences with and perceptions of smartphones and tablets, and an experiment involving controlled tasks on smartphones and tablets to gather interaction characteristics. By creating our logging application that captured detailed interaction logs while mimicking the appearance and functionality of actual application interfaces, we were able to observe realistic issues that the participants encountered, while also being able to visualize and analyze the details of their interaction styles and the difficulties they encountered.

Our observations revealed a number of specific issues and challenges encountered by our participants. For instance, a number of them experienced unintentional taps due to parts of their hand accidentally making contact with the touch screen, as well as unregistered taps due most likely to their fingers being dry. Our log visualizations also revealed specific interaction behaviors, such as differences in panning and zooming strategies on the map application, with some users not utilizing zooming much at all and simply panning long distances to find the desired target.

Based on these and other observations from our study, we provided some design considerations for enhancing the three applications as well as suggested methods to train and guide the user towards gaining a better understanding of the issues they may encounter and how to deal with them.

Our exploration is preliminary, but the approach we adopted of gathering detailed logs while users performed tasks on realistic interfaces can yield many interesting and insightful findings. Since "elderly users" is a very broad sweeping category of users, no one study can comprehensively cover a representative collection of users from the category. Therefore, there is a need for a greater number of studies to be conducted looking at specific issues encountered by a wide variety of users. Further work also needs to be done to explore a wider range of applications and interaction methods, to explore other ways of effectively visualizing the interaction logs, and to implement and validate the proposed enhancements.

References

1. Plaza, I., Martín, L., Martin, S., Medrano, C.: Mobile applications in an aging society: Status and trends. Journal of Systems and Software 84, 1977–1988 (2011)
2. Docampo Rama, M., Ridder, H.D., Bouma, H.: Technology generation and age in using layered user interfaces. Gerontechnology 1 (2001)
3. Leung, R., Tang, C., Haddad, S., Mcgrenere, J., Graf, P., Ingriany, V.: How older adults learn to use mobile devices: Survey and field investigations. ACM Trans. Access. Comput. 4, 11:1–11:33 (2012)

4. Kurniawan, S.: Older people and mobile phones: A multi-method investigation. Int. J. Hum.-Comput. Stud. 66, 889–901 (2008)
5. Lee, Y.S.: Older adults' user experiences with mobile phones: Identification of user clusters and user requirements (2007), `http://scholar.lib.vt.edu/theses/available/etd-09172007-135013/`
6. Renaud, K., Van Biljon, J.: Worth-centred mobile phone design for older users. Univers. Access Inf. Soc. 9, 387–403 (2010)
7. Van Dyk, T., Renaud, K., Van Biljon, J.: Moses – method for selecting senior mobile phones: Supporting design & choice for the elderly. In: Proceedings of the South African Institute for Computer Scientists and Information Technologists Conference, pp. 277–285. ACM, New York (2012)
8. Jastrzembski, T.S., Charness, N.: The model human processor and the older adult: Parameter estimation and validation within a mobile phone task. J. Exp. Psychol. Appl. 13, 224–248 (2007)
9. Kobayashi, M., Hiyama, A., Miura, T., Asakawa, C., Hirose, M., Ifukube, T.: Elderly user evaluation of mobile touchscreen interactions. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) INTERACT 2011, Part I. LNCS, vol. 6946, pp. 83–99. Springer, Heidelberg (2011)
10. Leitão, R., Silva, P.A.: Target and spacing sizes for smartphone user interfaces for older adults. hillside.net. (2012)
11. Nicolau, H., Jorge, J.: Elderly text-entry performance on touchscreens. In: Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility, pp. 127–134. ACM, New York (2012)
12. Wacharamanotham, C., Hurtmanns, J., Mertens, A., Kronenbuerger, M., Schlick, C., Borchers, J.: Evaluating swabbing: A touchscreen input method for elderly users with tremor. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 623–626. ACM, New York (2011)
13. Ziefle, M., Bay, S.: How older adults meet complexity: Aging effects on the usability of different mobile phones. Behaviour & Information Technology 24, 375–389 (2005)
14. Caprani, N., O'Connor, N.E., Gurri, C.: Touch screens for the older user. In: Auat Cheein, F. (ed.) Assistive Technologies. InTech (2012)
15. Czaja, S.J., Charness, N., Fisk, A.D., Hertzog, C., Nair, S.N., Rogers, W.A., Sharit, J.: Factors predicting the use of technology: findings from the Center for Research and Education on Aging and Technology Enhancement (CREATE). Psychol. Aging. 21, 333–352 (2006)
16. Pattison, M., Stedmon, A.: Inclusive design and human factors: Designing mobile phones for older users. PsychNology Journal 4, 267–284 (2006)
17. Morris, A., Goodman, J., Brading, H.: Internet use and non-use: Views of older users. Univers. Access Inf. Soc. 6, 43–57 (2007)
18. Zhou, J., Rau, P.-L.P., Salvendy, G.: Use and design of handheld computers for older adults: A review and appraisal. International Journal of Human-Computer Interaction 28, 799–826 (2012)
19. Weilenmann, A.: Learning to text: An interaction analytic study of how seniors learn to enter text on mobile phones. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1135–1144. ACM, New York (2010)
20. Leonardi, C., Albertini, A., Pianesi, F., Zancanaro, M.: An exploratory study of a touch-based gestural interface for elderly. In: Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries, pp. 845–850. ACM, New York (2010)

21. Lepicard, G., Vigouroux, N.: Touch screen user interfaces for older subjects: Effect of the targets number and the two hands use. In: Miesenberger, K., Klaus, J., Zagler, W., Karshmer, A. (eds.) ICCHP 2010, Part II. LNCS, vol. 6180, pp. 592–599. Springer, Heidelberg (2010)
22. Caprani, N., Doherty, A.R., Lee, H., Smeaton, A.F., O'Connor, N.E., Gurrin, C.: Designing a touch-screen sensecam browser to support an aging population. In: CHI 2010 Extended Abstracts on Human Factors in Computing Systems, pp. 4291–4296. ACM, New York (2010)
23. Tsai, T.-H., Chang, H.-T., Wong, A.M.-K., Wu, T.-F.: Connecting communities: Designing a social media platform for older adults living in a senior village. In: Stephanidis, C. (ed.) Universal Access in HCI, Part II, HCII 2011. LNCS, vol. 6766, pp. 224–233. Springer, Heidelberg (2011)
24. Sato, D., Kobayashi, M., Takagi, H., Asakawa, C., Tanaka, J.: How voice augmentation supports elderly web users. In: The Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility, pp. 155–162. ACM, New York (2011)

From Persona to Techsona

Susanne Bødker and Clemens Nylandsted Klokmose

Department of Computer Science, Aarhus University
{bodker,clemens}@cs.au.dk

Abstract. In this paper we introduce the notion of *techsona*, as a possible counterpart design instrument to personas. We use a case study to illustrate a design process where techsonas help pinpoint technological ideas and alternatives and analyze these systematically. While a persona captures "imagine a user…", a scenario "imagine a situation…" the techsona gives the missing piece of "imagine a technology…" We show how a recently developed activity theoretical model can serve both as a framing for techsonas and as an analytical interface between personas and techsonas. The paper discusses the potentials and problems of the techsona and concludes that the techsona truly helps with a missing piece in persona and scenario-based design.

Keywords: Persona, Techsona, Activity Theory.

1 Introduction

In our work with interaction design, we have used scenarios and personas in various forms over the years. The capacity, and possible and plausible actions of humans are captured and communicated through these. Hence, the use of personas and scenarios help us answer the question of how potential users reach out towards current or hypothetical future technologies. The reverse side, however, needs to be addressed similarly: How do the action possibilities of an artifact, whether a current or hypothetical artifact, reach out towards the potential user? We miss a set of design instruments, matching personas, in order to capture and communicate open issues regarding the artifacts designed, design decisions and even technological alternatives. In this paper we present *techsonas* to provide this match.

Personas are connected to scenarios in that they often work together, and share the same qualities of capturing both the current situation and human action possibilities on the one hand, and the future on the other [3,5]. Scenarios focus on particular situations, whether typical or critical, and they can generally be used throughout the design process to hold on to design decisions while being open for interpretation. It is our claim in this paper that we need the techsona to counterpart personas and scenarios in order to help users and designers reason about technology on a more abstract and hypothetical level than prototypes. We need means to capture and concretize technological alternatives and thought experiments without necessarily manifesting them as prototypes. Hence, we want to provide a light-weight yet systematic means for addressing questions like: *What if we used a wall-mounted multi-touch screen for our ticket vending machine, would that benefit or impede our users?*

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 342–349, 2013.

Accordingly, we need to populate the gap that we are addressing here with a tool that helps focus on the critical detail in terms of use of future artifacts, and acts as counterparts to personas at the levels from human fear and values to specific action possibilities and constraints. In addition this tool needs to capture elements of existing technologies that are useful when designing the future.

2 From Personas and Scenarios to Techsonas and Back Again

Personas [11, 12, 13] are defined as ways of representing real people throughout the design process. According to Cooper [13], personas are hypothetical archetypes, defined with rigor and precision, discovered in the investigation process. Personas allow for an understanding of the values, fears, etc., of the users [13]. Personas are used in two ways: by their designers in their own design process and for communication with others [17]. In interaction design personas are often used in conjunction with scenarios. Scenarios are most well known from Carroll [9]. He points out that scenarios are stories about people and their activities; that they are specific and fluid at the same time and they point towards both analyses of existing use/work activities and visioning of the future. Making scenarios is a creative process: they are hypotheses, or qualified guesses about the artifact and its use [5]. They serve to open the dialogue about future possibilities and current constraints. [4, 5] propose to work from use-like situations towards scenarios that are constructions meant to stage acting in the future or to reflect on and illustrate problems with this action. Our use of personas and scenarios in design is described in [8].

While Carroll [9] argues *"…that use-scenarios can be the principle design representation of an artifact"* (pp. 190) Bødker & Christiansen [5] propose to capture and communicate innovative or prototypical examples of technologies, in a manner that seems quite parallel to personas. It is necessary in particular to enable a systematical exploration of new ideas, which is an integral part of any type of interaction design. Hence, we need a means for exploring an (even far-fetched) idea against personas and use scenarios in a lightweight yet systematic manner. Accordingly we see personas and techsonas as closely connected with scenario-based methods, and we have been inspired by previous work on future scenarios [3] and inspiration cards [16].

The difference between personas, techsonas and scenarios can briefly be captured as follows. While a persona in design would easily be approached as "imagine a user…", a scenario as "imagine a situation…" the techsona gives the missing piece of "imagine a technology…". If focusing on a current use situation, personas summarize and crystalize findings regarding users' actions and motives, scenarios flesh out (specific) use situations, and techsonas summarize the action possibilities, or affordances of the technology. Like a persona, the techsona is hypothetical, yet rigorous and precise enough to be confronted with a variety of personas and scenarios.

Where personas are discovered through investigation, techsonas are invented as part of the design process. A good techsona embodies a simple (potential) design idea and provides enough description of its realization to confront the values, fears, capabilities, etc. of the users as expressed in the personas.

In the following we discuss what goes into a techsona and how can it be confronted with personas in a systematic manner. We use a model derived from activity theory that helps reasoning about the dialectics between human and technology.

2.1 A Framework for Mediation between Personas and Techsonas

Recently we have applied the Human-Artifact Model [7] for design-oriented analyses of current and future artifacts and activities of use. In this context the model provides an emphasis on understanding interaction with technology as multi-layered and dialectical. The analytical scheme of the Human-Artifact Model (Figure 1) combines analyses of human experiences and artifacts, and addresses the tensions between human skills and capacity on the one hand, and the action possibilities and affordances offered by the artifact on the other [3]. This is done on three levels reflecting the activity hierarchy: activity, action and operation. These levels provide three sets of analytical glasses: Motivation (by asking why?), goal-orientation (by asking what?) and operation (by asking how?).

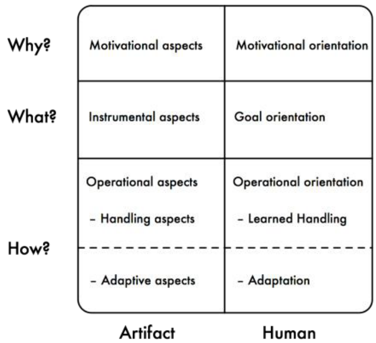


Fig. 1. The Human-Artifact Model (according to [7])

We have combined this approach to understanding the relationship between a practice and given artifacts with iteratively and concurrently developing personas [8]. Through this we confront the richness of personas with the more focused questions of "Why?", "What?" and "How?" coming from the Human-Artifact Model (Figure 1). In the following we illustrate how this is done similarly for the techsona. The model resembles Carroll et al's task-artifact cycle [10. However, while the task-artifact cycle reflects a causal relationship between the artifact and task, the dialectical relationship of the Human-Artifact Model is different: In practical terms, we need to analyze what happens and might happen to the human activity every time artifacts are changed and vice versa, but there no causal predictions available, such as 'because the artifact is changed in this way, the task will change in that'. Another difference between the Human-Artifact Model and the task-artifact model in terms of its focus on tasks, in short, is that tasks are mainly about actions, whereas the *why* level in particular is missing in the analysis. We refer to [1] for a longer discussion of tasks.

Applying the Human-Artifact Model as a mediator between personas and techsonas emphasizes their dialectical relationship, and provides a structure for systematically exploring the tensions between the assumptions of the artifact and the capabilities and orientation of the user towards the artifact all the way from motivation to low-level operation.

3 Case: The Municipal E-service

The eGov+ project focused on studying and improving eServices for public authorities (see e.g. [6]). The main problem identified were that public authorities often see eServices as replacements of the ways in which citizens show up in the public office to receive e.g. a new passport or a driver's license. The project worked to identify alternative ways of designing and understanding citizen eServices and multiple student projects have been part of this.

Based on records of interviews and observations with citizens and municipal workers, the researchers developed a set of personas that the students were asked to use as starting points when carrying out an iterative design process (see more in [8]). Personas were used to summarize substantial research findings as communication devices within the project, from researchers/analysts to students/designers, similar to the ways personas is described in [11, 15]. They were supplemented with scenarios to describe the current use situation as well as future use. Here is a condensed description of the persona Ahmed (from [6]):

"Ahmed is a citizen of the world. Yet, somewhat in contrast to this, he is motivated in his life by a wish to own his own home, to live in Denmark. He prefers face-to-face deals in general, and he does e.g. most of his shopping in this manner. He is heavily dominated by his concerns for his close family and their everyday life in Denmark (school, housing, etc.) and his wider family. This family is spread across the world and he makes contact with members whenever possible. Ahmed is used to visit the municipal office in question. He is used to cell phones whereas he does not use Internet banking. He travels with his family to meet relatives whenever possible."

Based on a more detailed description of Ahmed and five other personas, and a selection of scenarios [8] the students were asked to discuss technology suggestions, such as a passport machine coming from a newspaper article, to make rich descriptions of this and their own technological alternatives and relate them to the given personas–these rich descriptions are what we refer to as *techsonas*. A condensed example the techsona "friend-in-waiting": *"The Friend-in-waiting allows citizens who come to the citizen services office to spend the time in the waiting area completing as much passport information as possible. It is used in a similar way as current paper forms, but has the advantages that existing information from municipal systems can be called up automatically. The Friend-in-Waiting is a pad that the front-desk staff hands to people when they enter the office. Hence the Friend-in-waiting does not require for citizens to bring their own device, if only they bring their social security card. The look-and-feel resemblance with paper forms combined with the use of touch* *displays is intended to help citizens who are otherwise not using computer technology. The Friend-in-Waiting is also a guide that leads citizens to a booth where pictures are taken and automatically linked to the ID of the citizen, and, if needed, to the citizen services staff responsible for handling the passport."*

The students used The Human-Artifact Model as an interface between personas and techsonas, to analyze potential matches and mismatches between human beings and a technology as it was described, or tensions between persona and techsona. The Human-Artifact Model was also used to structure the development of a techsona, e.g. when the students were asked to develop a techsona based on the passport machine described in a newspaper (Figure 2).

Fig. 2. Creation of techsona through answering the questions of the Human-Artifact Model (The Passport Machine as read about in the newspaper)

From there the students directed their attention towards the future and used the given personas personas to develop scenarios of use, and the Human-Artifact Model to structure their insights regarding the users and use situations. Since the goal was to inform design of a new technological solution used when acquiring a new passport, the personas were coded to focus on e.g. Ahmed's orientation towards a potential new artifact. The Human-Artifact Model supported this coding through its levels: Why would he use the artifact? What will he need to do? How is he used to achieve the goals the artifact realizes? Whether such artifact was useful was debatable, and up for the students to uncover.

Along the way, the students were asked to be skeptic towards their ideas and to provide alternatives. Hence the outcome of the process was four sets of *alternatives* illustrating four different sets of problems relating to a passport machine. These included the possible use of mobile technology while in the municipal office or away from it. One of these alternatives was the ‘Friend-in-waiting’ – a tablet-based passport device to be used while in the waiting area of citizen services. This helped address a critical issue that had been uncovered in discussions of the passport machine, namely the match of the passport picture with the actual person, which, according to legislation, required physical presence at the office. But how would it help citizens like Ahmed? A techsona was created together with the video-based scenario and coded with the Human-Artifact Model for analysis towards the personas. This techsona was confronted with the personas and new issues occurred that were explored further through traditional means such as a video scenario (exemplified in Figure 3).

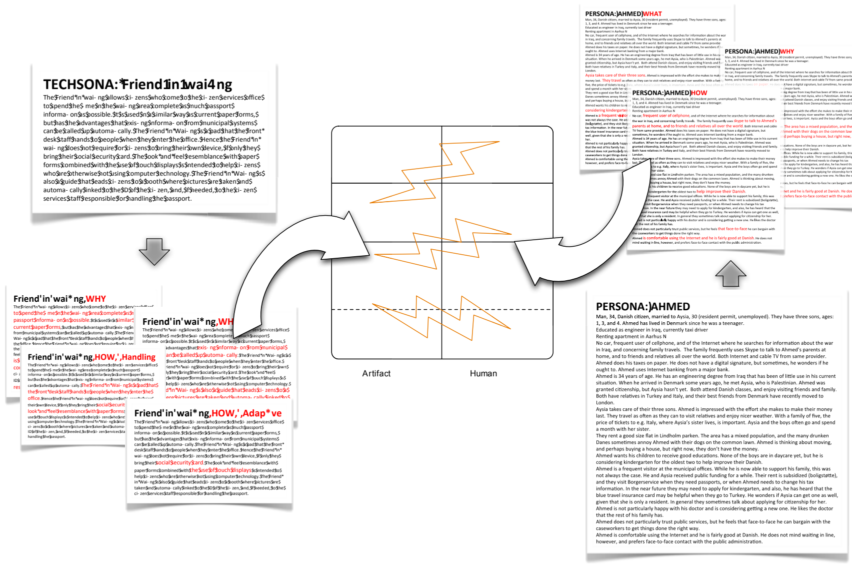


Fig. 3. Human-Artifact Model as a mediator between persona and techsona

4 The Meeting of Persona and Techsona

Underlying this work to establish and code the personas and techsonas lies the activity theoretical idea that the human users possess a set of learned and adapted action possibilities, and they are oriented towards certain goals and motives. All of these change throughout the human activity, and they are changeable when meeting and using a new artifact. A new artifact provides new possibilities for use, but it also talks back and cannot be used in any odd way the user may want. Similarly the user cannot necessarily make her capacities work with any new artifact.

Accordingly the meeting of the persona and the techsona is intended to help address the possibilities of a match between the human and the artifact side, as well as the problems. This is not an either/or, rather they are dialectical tensions: Where lie the possibilities that the Persona Ahmed (or one of the other personas) would use the ‘Friend-in-Waiting’, or the passport machine? What kinds of immediate problems do we find that may prevent such future use, or cause a need to adjust the concept before we move on to designing a more elaborate prototype? For instance, would Ahmed's preference for a personal handshake and face-to-face communication prevent him from picking up the 'Friend-in-Waiting'? Or is it sufficient that he meets a person when picking it up and leaving it? Would experience with modern multi-touch interfaces on mobile devices help or hinder his handling of the pad? Or would the idea of picking up a device that is not his own be a barrier? etc. (Figure 3).

The sort of analysis illustrated in this example falls in a tradition that goes back to [4], where the person side and the technology side of activity are systematically confronted with each other through a number of different scenarios, based on systematic questions. The Human-Artifact Model has helped making it more operational in terms of helping students and other designers work in a structured manner with the levels and questions. What we are providing is an analytical tool that helps uncover potentials and problems of future use at an early stage, but obviously, ultimately, there are no firm answers, and this sort of analysis does not prevent problems from occurring in later prototypes or the final design.

5 Discussion and Conclusion

In this paper we have introduced techsonas to facilitate a focus on technological possibilities and alternatives in scenario- and persona-based design. We have used the Human-Artifact Model as a mediator between these. The above example illustrates how the analysis brings out critical details, such as in the example, the relationship between experiences with personal devices related to a shared tablet in the case of "friend-in-waiting". While we have limited evidence for this, we suggest that through this kind of analysis we may be able to recognize important details from less important ones, in techsonas, as well as in the personas.

In line with the original motivation for personas, the techsona makes it possible to address motivational issues together with the handling possibilities of the proposed technology. Hence, the techsona targets the role of proposed technology in relation to e.g. human fear and values, and the techsona in that manner is a way of filling the gap that we identified.

There may well be other ways of filling this gap, and we suspect that the Human-Artifact Model structured analysis is best suited for the design of IT, and less so for e.g. a media production where personas similarly are used to represent the target audience. Other theoretical frameworks may serve equally well as a mediator between techsonas on the one hand and scenarios and personas on the other hand. However, the activity theoretical rooting in dialectical thinking is in our opinion crucial to properly understand the relationship between personas and techsonas.

In our experiences with techsonas so far, we have seen indications that this analysis to some extent helps resisting the temptation of premature prototyping, i.e. the temptation to pursue and build a prototypes simply because it looks like a good idea, technically. The students who carefully used the techsonas seemed more critical to their own ideas. Obviously we have built our presentation around one, simplified example.

Richer material and more cases are behind this presentation, and we are working to gather even more such material. In the meantime, we encourage the readers to start using and developing the ideas further.

Acknowledgements. We thank members of the eGov+ project as well as Karen Johanne Kortbek, Lene Mejlby, Nikolaj G. Borchorst, Marianne Dammand Iversen, Peter Dalsgaard and Kim Halskov for their various contributions. This work has been supported by the Carlsberg Foundation and Aarhus University through the iCenter PIT.

References

1. Bannon, L., Bødker, S.: Beyond the Interface, Encountering Artifacts in Use. In: Carroll, J. (ed.) Designing Interaction: Psychological Theory of the Human-Computer Interface, pp. 227–253. CUP (1991)
2. Beguin, P.: In search of a unit of analysis for designing instruments. Artifacts 1(1), 12–16 (2007)
3. Bødker, S.: Scenarios - setting the stage for reflection and action in user-centered design. Interacting with Computers 13(1), 61–77 (2000)
4. Bødker, S., Christiansen, E., Thüring, M.: A conceptual toolbox for designing CSCW applications. In: COOP 1995, Juan-les-Pins, pp. 266–284 (January 1995)
5. Bødker, S., Christiansen, E.: Scenarios as springboards in design. In: Bowker, G., Gasser, L., Star, S.L., Turner, W. (eds.) Social Science Research, Technical Systems and Cooperative Work, pp. 217–234. Erlbaum (1997)
6. Bødker, S., Christiansen, E., Nyvang, T., Zander, P.O.: Personas, people and participation–challenges from the trenches of local government. In: 12th Participatory Design Conference, pp. 91–100 (2012)
7. Bødker, S., Klokmose, C.N.: The Human-Artifact Model: An Activity Theoretical Approach to Artifact Ecologies Human-Computer Interaction. HCI Journal, 315–371 (2011)
8. Bødker, S., Klokmose, C.N.: Preparing students for (inter-)action. Journal of Design (2012)
9. Carroll, J.M.: Making Use: Scenario-Based Design of Human-Computer Interactions. MIT Press (2000)
10. Carroll, J.M., Kellogg, W.A., Rosson, M.B.: The Task-Artifact Cycle. In: Carroll, J.M. (ed.) Designing Interaction: Psychology at the Human-Computer Interface. CUP (1991)
11. Chang, Y., Lim, Y., Stolterman, E.: Personas: from theory to practices. In: Nordi-CHI 2005, pp. 439–442. ACM (2008)
12. Cooper, A.: The Inmates are Running the Asylum. SAMS (1999)
13. Friess, E.: Personas and decision making in the design process: An ethnographic study. In: CHI 2012, pp. 1209–1218. ACM (2012)
14. Grudin, J., Pruitt, J.: Personas, participatory design, and product development: An infrastructure for engagement. In: PDC 2002, pp. 144–161 (2002)
15. Halskov, K., Dalsgaard, P.: The emergence of ideas: the interplay between sources of inspiration and emerging design concepts. CoDesign 3(4) (2007)
16. Matthews, T., Judge, T.K., Whittaker, S.: How do designers and user experience professionals actually perceive and use personas? In: CHI 2012, pp. 1219–1228. ACM (2012)

Going Global with Personas

Lene Nielsen[1], Kira Storgaard Nielsen[1], Jan Stage[2], and Jane Billestrup[2]

[1] IT University of Copenhagen, Rued Langgaardsvej 7, 2300 Copenhagen S, Denmark
[2] Aalborg University, Selma Lagerlöfs Vej 300, 9220 Aalborg East, Denmark
{lene,kist}@itu.dk,
{jans,jane}@cs.aau.dk

Abstract. The persona method is widely used and commonly described both in scientific literature and in case-based blogs. Most often the descriptions point to a local context with local user groups and it is difficult to find writings on use of the method in an international context and in globally distributed teams. This paper reports from a qualitative study conducted in 2012/13 within 13 Danish companies and points to how design teams apply several different strategies when end-users are distributed worldwide. Moreover it shows how the designers value the strength of the method to provide common grounds for the team, especially for team distributed across countries.

Keywords: personas, scenarios, cross culture, international, design.

1 Introduction

Personas are descriptions of fictitious users and it is recommended that they get constructed from different forms of field data [1], such as surveys, user interviews, observations, and a combination hereof. The method enables empathy and engagement in the end-users [2] and helps designers to imagine the persona in a future use situation. Personas are used for different design activities that are typically explored in stories – scenarios - that describe future possibilities in an easily accessible way and in a format that is easy to change. The persona method has developed from being a method for IT system development to being applied in many other contexts, including development of products, marketing, communication planning, and service design.

The persona descriptions have over the years developed a set layout. The persona has a name, the description is most often 1-2 pages long, and has several subsections that describe the persona's characteristics, life, behaviors, and preferences. Most often a photo accompanies the description [3].

The perceived benefits of personas for design range from increasing the focus on users and their needs, being an effective communication tool, to having direct design influence, such as leading to better design decisions and defining the product's feature set [4-9]. The method is criticized for: being too founded on qualitative data and non-scientific, being difficult to implement, not being able to describe actual people as it only describes characteristics, and for preventing designers meeting actual users [1].

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 350–357, 2013.

Despite criticism the common perceived benefits are two-fold: when designing products the method facilitates that designers remember that they differ from the end-users, and that the method enables designers to envision the end-user's needs and wants. Furthermore the method creates empathic understandings of users and breaks the designers' automated perceptions of end-users, which are grounded in individual experiences and not in data. Finally the persona method supports a common and aligned understanding of the end-users that is shared among project teams.

Fig. 1. An example of a typical persona description from The Danish Tax Authorities - Skat.dk. In this example the borders are color coded in order to heighten the perception of differences among the compilation of personas.

Most persona descriptions are to be used for country specific projects, but international companies also use the method in projects that are aimed at an international audience and used by globally distributed design teams. In this paper we investigate the strategies applied when Danish companies develop personas for an international audience, the outcomes, and the perceived benefits and difficulties.

Creating personas for an international audience is not well described in the literature. Putnam et. al. [10] describes two cases of conveying data for personas from Kyrgyzstan and the Andhra Pradesh region of India and the strategies they employed for the persona descriptions that were to be used by designers, developers, and other stakeholders not from these regions. One of the strategies was to use scenarios in the persona descriptions to convey cultural and lifestyle differences.

Snyder et. al. [11] describe a case where they discussed three approaches to integrating cultural differences into persona descriptions: 1) A separate persona for each culture and for each task. 2) U.S. based persona descriptions, each including sections with cultural differences. 3) One persona from each country visited, with the cultural differences as part of the descriptions. In this case the team realized that there were few cultural differences. They ended up with persona descriptions from several countries as a reminder to the team that the product is used in different countries and cultures.

Finally we identified a case description by Windows International Program Manager's team in Pruitt and Adlin [9] where the team identified personas for each ethnographic region and used a scoring on how different these were on selected parameters from the US personas. The team then defined the differences in the persona descriptions in accordance to US.

2 The Study

This paper reports from a qualitative study with 28 participants from 13 Danish companies that are experiences in using the persona method. Their experiences range from 6 month to more than 10 years. The study took place from October 2012 to January 2013. The study focuses on the companies' use of the persona method. The interviews are analyzed for statements on cross-cultural issues such as international personas that cover several nationalities and the use of the method in globally distributed design teams.

2.1 Finding the Companies

In order to find the companies that have experience in using personas we used different online channels such as: personas.dk, Infinit.dk and Linkedin groups such as User Centric Network, Service Design in Denmark, and Sigchi.dk, we advertised for companies in two news letters respectively from Infinit and Sigchi.dk, and used word of mouth.

To broaden the perspective and get companies outside of our own network companies were found by search on keywords such as: personas and usability.

The intention was to cover public and private institutions, as well as large, middle-sized and small companies.

In the end we did interviews at public institutions: Danish Broadcast, The Royal Danish Library, The Danish Tax Authorities, Aarhus Libraries. Large private companies: Danske Bank, Microsoft Solutions, Safecom, and Widex. Middle sized private companies: Mjølner Informatics, AdviceDigital. And small private companies: Centre for Digital Pedagogy, Value-Creating Construction, and The Food Culture Zone.

2.2 Method of Analysis

The participants were asked to describe their processes of work: how they collect and use data, and how they design and use the personas.

In line with Kvale's [12] advice to start the analysis already when interviewing we began identifying the recurring issues of working with personas in a global setting already in the interview situations. Furthermore we recorded the interviews in the audio transcription tool F4 and while listening to the interviews, passages focusing on use of personas in an international or global perspective, were identified through the method of condensation [12]. The condensed transcripts have been analyzed, first for statements on use and use situations, secondly for statements on global issues.

Then the statements were mapped into categories, in order to identify which benefits, challenges, and problems the companies face when using personas within an international setting. Moreover which strategies they use to overcome the challenges.

As this study is part of a larger project about the use of personas in general, it has been possible in all phases of the analysis to compare the national and international perspective thus to identify the additional problems and benefits of the method when companies face an international user group.

3 Findings

3.1 The Value of Personas

In general the benefits of using personas for design are viewed from a strategic level and with a production perspective.

> C11: "We are still quite technically oriented and quite nerdy when we develop things. And now it's the customer's needs that are described first. I'm not saying this happens in all the cases, but it is a step towards it. This is totally different from what we did before. And personas has helped to understand what kind of needs it is you have to cover."

On a strategic level the benefits are:

- It provides guidance in the development process as it enables a focus on user needs instead of project participants' likings.
- The focus on user needs improves the definition of the project.
- In the decision making process the scenarios creates a clear understanding of what the users need and what goals the company wants to achieve.
- In the discussions the project participants' implicit understandings becomes explicit.
- A participant from IT consultancy mentions that they use the method to confront their clients with their lack of visions.

From a production perspective the benefits are:

- When working agile it helps prioritizing user stories[1].
- The method creates an understanding of how the product creates value for the end-users.
- The personas are used to create use cases and scenarios.

---

[1] User stories are short descriptions of functionalities [13].

3.2 The Value of Persona in an International Perspective

The benefits of the method are perceived as being even bigger in the companies with a global audience and globally distributed project teams. The perceived values of the persona method are several. One benefit of the method mentioned by all participants is its ability to create a common language and a shared understanding of who the users are.

> C12: “One thing is for sure, it has provided a common language. It might sound a bit like a cliché, but it really is a common language. I get quite warm at heart when you sit in a meeting where you hardly know anybody, and then somebody says: "Well, would Alex think like this?" Then we all know whom we're talking about, we know the audience, and whom we address - just because of the name. It is a common language and I think it is enormously important – this goes across our department and reaches out to our subsidiary companies."

For large corporations the benefits are twofold: a shared language to discuss and talk about users that are nationally and culturally diverse, and a shared understanding across departments and national borders.

3.3 Data

The participants from companies that have an international audience describe it as especially challenging to get enough data and to know when there is enough data. They try to solve this by collecting data from different parts of the world, but none of the participants have data from all of the markets they address. Thus the major challenges become to get enough data, to get the right data, and to know when there is enough data to cover all the important aspects.

In particular it is mentioned that there is a difference making personas for a Western context and an Asian. Data on Asian users are hard to come by as this knowledge is not easy accessible and it is expensive to get data from countries were the Danish companies do not have any offices. Furthermore, from a Danish perspective, it is experienced that the cultural differences between the Danish and the Asian users are larger than those between Danish and European users. This in common makes it difficult to develop personas for Asian users.

3.4 Persona Descriptions

The companies who have users from different countries experience it as especially challenging to create persona descriptions. They often look at specific national differences that are connected to:

- IT knowledge and use,
- Internet access,
- Language skills,

- Level of education,
- Different levels of autonomy in the companies and how management is distributed.

3.5 Strategies for the Descriptions

There are mainly two strategies to address the challenges of a global target group: One strategy is to develop persona descriptions that represent users from different countries (e.g. one Finnish, one Polish, etc.), sometimes with the additional information that some personas may be more prevalent in some regions than in others. Another strategy is to have descriptions that are so general that they can fit a wide variety of countries. Finally some companies try to combine the two strategies, by dividing the persona descriptions into an overall part and a part that singles out national differences. In the descriptions the global perspective can be observed in both naming and the connected images.

> C8: "She is from Frankfurt, one is from Poland and one is from Norway. [...] We have just said to ourselves in order to create a broad representation of users and contacts we must have some personas from around the world. Whereupon we asked the question to our reference group: "Do we need personas from around the world?" They certainly thought so. They said it makes good sense as we sell worldwide. [...] It is not randomly selected countries, they are chosen to represent several countries."

3.6 Strategies for Naming

Names are internationalized to cover a broad international group. There seems to be a common understanding that if the name is English; such as Anna, Elizabeth, or Robert, it can refer to any culture or nationality, regardless of an Eastern or Western context. When using the strategy of heightening the national awareness the personas are given specific national names, in order for the reader to deduce the nationality from the name, e.g. a concrete Finnish name such as Pekka.

3.7 Strategies for Images

As with the names the images that accompany the persona description follow the strategies either to point to a specific nationality or to suppress identification of a specific nationality. For the first strategy the photos represent different types, e.g. Asians, Africans, and Europeans. In order to suppress national characteristics the persona illustrations are typically of Western appearance, or they are intentionally blurred or drawings with no national and typological characteristics.

One company chose to not have illustrations at all. According to the participants the company had discussed and agreed upon that a photo of a person from a specific country would hinder empathy and create a distance to the persona, if the project member were not from the same country. The same company had interestingly enough decided to create personas where the name and the description had specific national characteristics. They did however not consider that a text would hinder empathy in the way that a photo will.

3.8 Lack of Experience

Based on internal considerations and discussions each company develops their own strategy, but does not have any research on if and why the strategy works. They are aware of the fact that they lack evidence of the usefulness of the applied strategy both from company cases and research.

4 Conclusion

There is no doubt that companies with diverse international customers and companies with globally distributed teams perceive the use of personas as beneficial. The experienced benefits are larger than when the companies operate solely in a national context.

The benefits are two-fold: as a common point of reference for the many different departments and as a common understanding of the diverse user groups. For the last mentioned benefit the method furthermore prevents project participants to use own local experiences on the international audience.

In line with the sparse literature we saw several strategies applied in the descriptions 1) to create descriptions that could fit all nationalities (read: Western cultures) 2) to create country specific descriptions, 3) to create descriptions that fit all nationalities, but with additional culture specific information.

Even though the companies perceive personas for an international audience as beneficial, the described strategies hold difficulties. The strategy to create descriptions that fit all might result in superficial descriptions, this both because of sparse knowledge of all international user groups and the attempt to converse the diversity in the user groups. When coupling nationality with persona description the risk is that the focus on nationality might not be comprehensive and the differences among the personas are perceived as national, even though the differences might lie elsewhere. For the strategy of creating descriptions that fit all, but with additional culture specific information, the descriptions run the risk of being very long and confusing.

We found that there are great benefits of the persona method, but the understanding of the different strategies applied still needs to be investigated further.

The sparse literature on cross-cultural personas is of no help, it deals with communication of different national user groups to designers that might be distant from the end-users, but none includes distributed design teams in the studies. Proctor et. al. [14] propose to look into the cross-cultural decision-making processes in the phases of design, adoption, use, and support. In line with this and to support the companies we need to further investigate: how data can be efficiently gathered and analyzed for cross-cultural personas and we need to look at how they are perceived and used in globally distributed teams. Finally we need to develop different methods to describe international personas.

Acknowledgement. We wish to thank Infinit for supporting the research.

References

1. Nielsen, L.: Personas - User Focused Design. Human-Computer Interaction. Springer (2012)
2. Nielsen, L.: Engaging Personas and Narrative Scenarios, vol. 17, PhD Series. Samfundslitteratur, Copenhagen (2004)
3. Long, F.: Real or Imaginary - the Effect of Using Personas in Product Design. In: IES Conference. Irish Ergonomics Review, Dublin (2009)
4. Cooper, A.: The Inmates Are Running the Asylum. SAMS, Indianapolis (1999)
5. Cooper, A., Reimann, R., Cronin, D.: About Face 3.0: The Essentials of Interaction Design. Wiley (2007)
6. Pruitt, J., Grudin, J.: Personas: Practice and Theory (2003), `http://research.microsoft.com/en-us/um/people/jgrudin/publications/personas/pruitt-grudin.pdf`
7. Ma, J., LeRouge, C.: Introducing User Profiles and Personas into Information Systems Development. In: AMCIS 2007, paper 237 (2007)
8. Miaskiewicza, T., Kozarb, K.A.: Personas and User-centered Design: How Can Personas Benefit Product Design Processes? In: Design Studies, vol. 32(5), pp. 417–430 (2011)
9. Pruitt, J., Adlin, T.: The Persona Lifecycle: Keeping People in Mind Throughout Product Design. Morgan Kaufmann, San Francisco (2006)
10. Putnam, C., Kolko, B., Wood, S.: Communicating About Users in ICTD: Leveraging HCI Personas. In: The Fifth International Conference on Information and Communication Technologies and Development, pp. 338–349. ACM, New York (2012)
11. Snyder, M., Sampanes, A., White, B.-K., Rampoldi-Hnilo, L.: Personas on the Move: Making Personas for Today's Mobile Workforce. In: Marcus, A. (ed.) Design, User Experience, and Usability, Pt II, HCII 2011. LNCS, vol. 6770, pp. 313–320. Springer, Heidelberg (2011)
12. Kvale, S.: InterView. Hans Reitzel, København (1997)
13. Schwaber, K., Beedle, M.: Agile Software Development with Scrum. Prentice Hall, Upper Saddle River (2005)
14. Proctor, R.W., Nof, S.Y., Yih, Y., Balasubramanian, P., Busemeyer, J.R., Carayon, P., Chiu, C., et al.: Understanding and Improving Cross-Cultural Decision Making in Design and Use of Digital Media: A Research Agenda. International Journal of Human-Computer Interaction 27(2), 151–190 (2011)

On Users' Preference on Localized vs. Latin-Based CAPTCHA Challenges

Christos Fidas[1,2] and Artemios G. Voyiatzis[3]

[1] Department of Computer Science, University of Cyprus, 1678 Nicosia, Cyprus
[2] Interactive Technologies Lab, HCI Group, Electrical and Computer Engineering Department University of Patras, 26504, Patras, Greece
[3] Industrial Systems Institute, Research Center "ATHENA" PSP building, Stadiou Str., 26504, Platani Patras, Greece
christos.fidas@cs.ucy.ac.cy, bogart@isi.gr

Abstract. A Completely Automated Public Turing test to tell Computers and Humans Apart (CAPTCHA) is a widely used security mechanism for constructing a high-confidence proof that the entity interacting with a remote service is actually a human being. Stimulated by the facts that: a) nowadays CAPTCHA challenges are solely based on the Latin alphabet, b) currently Internet population consists in its majority of non-native-English speakers and c) numerous web sites consist of exclusively localized content, we conducted an empirical study aiming to examine the effect of various factors on users' preference in solving localized vs. Latin-based text CAPTCHA challenges. The study embraced a between-subject design using a self-developed localized CAPTCHA mechanism, capable of producing text challenges based on the participants' native alphabet. A total of 384 non-native English speakers participated in the frame of the reported study which followed an ecological valid experimental design. Analysis of interaction results provides interesting insights which can be taken into consideration for designing more usable CAPTCHA mechanisms.

Keywords: CAPTCHA, Usability, Security, Native Language (non-Latin) systems, localized CAPTCHA.

1 Introduction

A Completely Automated Public Turing test to tell Computers and Humans Apart (CAPTCHA) is nowadays utilized by numerous application service providers in order to defend themselves against automated software agents whose purpose is to degrade the quality of a provided service, whether due to misuse or resource expenditure. CAPTCHA challenges are utilized worldwide by millions of users and thus, it becomes obvious that having usability flaws in the design of CAPTCHA mechanisms, results in unacceptable tradeoffs for the users in terms of time and money. Thus, considering usability aspects in designing CAPTCHA mechanisms becomes nowadays a necessity.

Nowadays, the vast majority of deployed CAPTCHA mechanisms rely on recognition of distorted text with audio as an alternative means for those with vision problems.

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 358–365, 2013.

Research on CAPTCHA design has received significant attention lately. There is a growing demand to enhance both the security and the usability aspects of CAPTCHA mechanisms, aiming to offer high security standards to application service providers and interaction transparency to actual users. Therefore, the design of secure and usable CAPTCHA mechanisms is considered to be a challenging endeavor [1].

There are two main research directions for improving usable security of CAPTCHA mechanisms. The first one is towards breaking and improving existing mechanisms, resulting in the development of design guidelines for increased security [2, 3]. Being focused on security, CAPTCHA mechanisms are currently designed as tasks which are not integrated smoothly in the primary user interaction flow and many attempts exists aiming to provide more usable mechanisms [4-14, 22].

Existing text-based CAPTCHA mechanisms produce challenges with characters solely from the Latin alphabet. This is a "one-size-fits-all" approach that may not provide a viable long-term solution or even a good usability tradeoff, considering that not all Internet users are familiar with this alphabet. This argument is further strengthened bearing in mind that the Internet population consists in its majority of non-native-English speakers[1] and numerous web sites consist of exclusively localized content.

In this context, it is necessary to explore alternative constructs that consider the diversity of Internet users in terms of conventions, cultural, and cognitive backgrounds; tools and resources; and contexts in which CAPTCHA challenges are required to be solved. Towards this direction, we aim to explore the applicability of *localized CAPTCHA* i.e., a system that presents challenges with characters, glyphs, or other symbols in users' native language and cultural context. In contrast, a *Latin-based* one presents challenges using characters only from the Latin alphabet i.e., lowercase and/or uppercase letters from "a" to "z".

A localized CAPTCHA adapted to users' contextual factors may actually *increase* the usability of a system, in terms of efficiency and effectiveness, while *not decreasing* the security level needed for protecting the system. In this context, the work presented in this paper is primarily driven by the need to apply User Centered Design (UCD) methodologies related to design and evaluation of CAPTCHA mechanisms and contributes towards this direction by investigating whether individual differences, focusing mainly on users' lingual characteristics, affect efficiency and effectiveness on interacting with text-based CAPTCHA mechanisms.

The rest of the paper is structured as follows: first, we review the related works. Subsequently, we describe the context of an empirical study, the methods, and the developed hypotheses. Then, we analyze and discuss the findings of the study. Finally, we summarize and outline the implications of the reported research.

2 Related Works

The role and involvement of the real users and their unique characteristics is getting the attention of the research community lately. Based on over a million collected responses, an overall success rate of 96.1% was observed on the ReCAPTCHA service

[1] http://www.internetworldstats.com/stats.htm

[15]. The service presents garbled words from English scanned books that OCR software failed to recognize. The authors noticed that "*non-English speakers seem to perform slightly worse than English speakers*"' based on statistically significant differences in success rates correlated with the origin of the respondent's IP address [16].

A large-scale study involving over 1,000 volunteers and 21 different text-based CAPTCHA systems revealed among other that "*non-native speakers of English take longer to solve CAPTCHAs and are less accurate on CAPTCHAs that include English words*" [17]. Users needed on average 9.8 seconds to solve a text challenge and between 12 and 25 seconds for an audio challenge. Overall, native-English speakers were substantially faster and slightly more accurate on solving the challenges. We note that the recruitment platform that was used in the study (Amazon Mechanical Turk) is currently available only in the US and thus, it might be the case of bilingual speakers. The study neither evaluated CAPTCHA challenges in languages other than English nor the performance of non-native English speakers.

As part of one accessibility studies, a localized CAPTCHA based on the Urdu language was developed for web-based applications [17]. Urdu is widely spoken in India. A small-scale evaluation of the system involved 50 Urdu-knowing users with no or elementary knowledge of English language. The study found that users were more accurate and solved faster challenges in Urdu rather than in English. Similar results were reported in a recent study with Arab participants [21].

Furthermore, a lab usability study of text-based English CAPTCHA involved 24 participants who spoke Chinese and learned English as a second language [18]. The authors observed that "*When reading a string of alphanumeric characters, the participants spoke the alphabets in English, but the numbers in Chinese.*" This resulted in erroneous responses as for example due to the similar-sounding number "one" in Chinese and letter "e" in English.

Another study investigated users' perceptions towards localized and Latin-based CAPTCHA challenges [19]. According to the perceptions of the participants, current implementations do not provide an acceptable tradeoff solution with regards to usability. Analysis of responses revealed that even experienced users had severe difficulties in coping with current CAPTCHA challenges: "*every other participant claimed to need two or more tries each time to solve a challenge.*" Furthermore, participants stated to be positively attributed with regards to localized CAPTCHA challenges, especially when visiting web sites with localized content, since their interaction flow tends to be more natural.

A follow-up work examined the relation between the cognitive style of the user and their preference and performance upon interacting with text- and picture-based challenges [20]. The results of the study revealed an existing main effect relating to users' cognitive styles (Verbalizer or Imager) and a respective performance issue in solving text- vs. picture-based CAPTCHA challenges.

We build on top of these findings aiming to investigate whether individual differences affect preference to localized- vs. Latin-based CAPTCHA challenges. We describe in the next section a study that utilized the developed localized CAPTCHA mechanism, the hypotheses that have been formulated for the purpose of our research together with analysis of interactions and the interpretation of results.

3 Method of Study

3.1 Procedure and Research Hypotheses

A research prototype of a localized CAPTCHA mechanism was designed and developed aiming to produce challenges of random characters based on different national alphabets. The research prototype was based on an open source CAPTCHA mechanism which has been enhanced as to produce CAPTCHA challenges based on UTF-8 and thus capable to produce challenges in a variety of national alphabets. The research prototype produced challenges in Latin and non-Latin alphabet with a similar level of text distortion. We formulated the following null hypotheses for the purpose of this research:

***H_0*.** There is no general preference of users towards Latin or non-Latin CAPTCHA challenges, considering also various main effects and interactions with respect to language presentation of the web site and users' national alphabet.

3.2 Experimental Design

The aim of the experiment was to examine user preference towards a specific type of CAPTCHA challenge by considering also various main effects and interactions with respect to language presentation of the web site and users' national alphabet. The participants were asked to visit a web page which was placed in front of a web application that announced the students' marks for the course assignments. In this study, the web page content appeared equal randomly in either local language (in Greek) or translated in the English language. The aim was to examine whether the web page content presentation language influenced the user preference towards selecting a specific type of CAPTCHA challenge.

Once a student initiated a session, a random process chose if the student would receive content in English or local language. This choice was fixed for the whole session. The page explained what a CAPTCHA challenge is and presented two example challenges, one in local language and one in English. It then asked the user to select the CAPTCHA type of preference. The placement of the two challenges was equally split: in half cases the localized CAPTCHA example challenge appeared on the left side of the screen and the English one on the right; in the other half this order was reversed. Once the CAPTCHA type choice was made, the participant was faced with the actual CAPTCHA challenge and was asked to provide the correct response. If a correct answer was received, the system displayed the marks for the course assignments or else it challenged again the participant with a new challenge in the same (previously selected) language of preference.

3.3 Demographics of the Participants and Validity of the Study

The participants of the experiment were juniors and seniors of two different academic institutions studying computing science and applications. They were all native non-English speakers. We recruited 384 students in total for this experiment. The age of the students was between 20 and 25 years old with only minor exceptions. In total, 440 different challenges were presented to these students.

The validity of a study is related to its internal and external validity, and ecological factors. There has been an effort to increase ecological validity of the research, since the participants were involved at their own working conditions and physical environments (without the intervention of any experimental equipment or person). Furthermore, participants were asked to solve the CAPTCHA challenge during a real-life task (the primary task of the participants has been to view their course marks). This closely resembles a typical usage scenario related to CAPTCHA usage in which the challenge is as a secondary task. In terms of external validity, we mention that further empirical research with a larger sample is needed aiming to compare the findings of this study with ones performed using other alphabets, more representative samples, and diverse cultural contexts.

4 Study Results and Discussion

In this section, we present and discuss the main findings of the study based on the analysis of the collected responses.

The experimental design explored the actual user preference and the effect of content presentation in this choice. The users visited a web page that contained an example of both a Latin-based and a localized CAPTCHA and asked them which language they prefer to be actually challenged. The page appeared either in local language or in English and the appearance order of the two examples (left-right) was random. For the purpose of the study, we separated the participants in two groups based on their skills with regards to the English language. The control Group (A) consisted of N=298 participants who were proficient in English. The experimental group (B) consisted of N=86 participants who were not proficient in English.

In Table 1, we summarize the CAPTCHA preferences for each group according to the users' choice. Those participants belonging to Group (A) select to solve the English challenge in the majority of cases (188/298), while those participants belonging to Group (B) select to solve the localized one (72/86). This difference in preference is significant $x^2(1,N=384)=58.650$, $p<0.001$.

Furthermore, we examined the effect of the language presentation of the web content towards users' preference for a specific type of CAPTCHA challenge. The choice of the participants in both groups is not affected by the page content language: control group $x^2(1,N=298)=0.220$, $p=0.639$ and experimental group $x^2(1,N=86)=0.332$, $p=0.564$. The choice is also not affected by the order in which the two example challenges appear: control group $x^2(1,N=286)=1.828$, $p=0.176$ and experimental group $x^2(1,N=86)=3.084$, $p=0.079$.

The aforementioned results suggest that there is a main factor between the user preference on CAPTCHA challenges and their familiarity with the Latin alphabet. It appears that, if the users are given a choice, then the users belonging to the experimental Group (B) prefer to solve challenges in their native language, whereas users belonging to the control group (A) prefer to solve the Latin-based CAPTCHA challenges.

Table 1. Group Preference on challenge language by page content

	Page Content		Total
	Latin	Localized	
	Group (A)		
English	87	54	141
Localized	101	56	157
Total	188	110	298
	Group (B)		
English	7	42	49
Localized	7	30	37
Total	14	72	86

The aforementioned results are based on objectively measured interaction data and are in line with the results presented in [19], which investigated users' preference in solving localized vs. Latin-based CAPTCHA challenges through a questionnaire based survey. Bearing in mind that these recent studies followed different methodological approaches (objective vs. subjective self-reporting data) it seems promising to consider user's diversity with regards to their lingual characteristics in the design of CAPTCHA challenges, as nowadays CAPTCHA designs and implementations clearly fail to consider in their iteration design cycles such individual characteristics of the users as they produce challenges with characters solely from the Latin alphabet.

Simultaneously, such an approach would have as well some positive side effects, especially in use cases in which the users interact with a web application in native language content presentation, as it is preferable to have a uniform presentation, including the actual site content and the CAPTCHA challenge. Thus, we argue that enhancing existing Latin-based CAPTCHA mechanisms with the capability to provide localized text-based challenges would result to a better user experience.

5 Conclusions

The purpose of this paper is to presented results of an empirical study with the aim to gain knowledge for enhancing usability and user experience related to CAPTCHA challenges as an effort to apply User Centered Design (UCD) approaches related to CAPTCHA mechanisms. We argue that for designing more usable mechanisms, research in this area should partially move its focus away from the technical security issues towards understanding the users and developing approaches which can be applied for increasing usability and offering a better user experience to users who are engaged in solving CAPTCHA challenges.

In this context, given that current CAPTCHA designs provide challenges solely in Latin alphabet and considering the diversity of World Wide Web users with regards to their lingual characteristics we developed a CAPTCHA mechanism capable to produce challenges of different alphabets. We utilized the developed CAPTCHA mechanism in the frame of an ecological valid experimental design aiming to examine users' preference towards Latin and localized CAPTCHA challenges. The analysis of results was mainly focused on users' preference towards localized and Latin-based CAPTCHA challenges. However, we also looked at the language presentation factor of the web page in which the CAPTCHA mechanism was embedded aiming to examine if this factor has a main effect towards users' preference to a specific type of CAPTCHA challenge. We recruited overall 384 non-native English-speaking users aiming to investigate the aforementioned factors and revealed that there is a statistical significant preference of users towards using specific type of CAPTCHA challenges depending on their profession in using the Latin-alphabet.

A future research direction, aiming to examine in wider perspective human factors on CAPTCHA designs, is to investigate preference and performance factors in localized and Latin- based CAPTCHA challenges in comparative cross-national studies. Bearing in mind that CAPTCHA tasks are not embedded smoothly into the primary task of a user (e.g. a user not familiar with the Latin alphabet performs a registration on a local content web site that requires to solve a Latin – based CAPTCHA challenge) and that every second millions of users world wide are solving CAPTCHA challenges, studies like the reported one could have an impact on making these tasks more affordable as even a small improvements will have massive benefits for users.

References

1. Yan, J., El Ahmad, A.S.: Usability of CAPTCHAs or usability issues in CAPTCHA design. In: Proceedings of ACM SOUPS 2008, pp. 44–52 (2008)
2. Burzstein, E., Martin, M., Mitchell, J.C.: Text-based CAPTCHA Strengths and Weaknesses. In: Proceedings of ACM CSS 2011, pp. 125–138 (2011)
3. El Ahmad, A.S., Yan, J., Tayara, M.: The Robustness of Google CAPTCHAs. Newcastle University Tech. Report CS-TR-1278 (September 2011)
4. Matthews, P., Zou, C.C.: Scene tagging: image-based CAPTCHA using image composition and object relationships. In: Proc. of the 5th ACM Symposium on Information, Computer and Communications Security (CCS 2005), pp. 345–350. ACM, New York (2005)
5. He, P., Sun, Y., Zheng, W., Wen, X.: Filtering short message spam of group sending using CAPTCHA. In: Proceedings of the First International Workshop on Knowledge Discovery and Data Mining (WKDD 2008), pp. 558–561 (2008)
6. Kluever, K.A., Zanibbi, R.: Balancing usability and security in a video CAPTCHA. In: Proceedings of ACM SOUPS 2005, pp. 14:1–14:11. ACM, New York (2005)
7. Zhang, H., Wen, X., He, P., Zheng, W.: Dealing with telephone fraud using CAPTCHA. In: Proceedings of the 2009 Eight IEEE/ACIS International Conference on Computer and Information Science, pp. 1096–1099. IEEE Computer Society, Washington, DC (2009)
8. Datta, R., Li, J., Wang, J.Z.: IMAGINATION: A robust image-based CAPTCHA generation system. In: Proceedings of the 13th Annual ACM International Conference on Multimedia, pp. 331–334. ACM, New York (2005)

9. Cui, J., Zhang, W., Peng, Y., Liang, Y., Xiao, B., Mei, J., Zhang, D., Wang, X.: A 3-layer dynamic CAPTCHA implementation. In: 2010 Second International Workshop on Education Technology and Computer Science (ETCS), pp. 23–26 (2010)
10. Lin, R., Huang, S.Y., Bell, G.B.: A new CAPTCHA interface design for mobile devices. In: Lutteroth, C., Shen, H. (eds.) Australasian User Interface Conference (AUIC 2011), pp. 3–8. ACS (2011)
11. Banday, M., Shah, N.: Image Fip CAPTCHA. ISeCure, The ISC International Journal of Information Security, 105–123 (2009)
12. Ho, C.J., Wu, C.C., Chen, K.T., Lei, C.L.: DevilTyper: a game for CAPTCHA usability evaluation. Comput. Entertain. 9, 3:1–3:14 (2011)
13. Holman, J., Lazar, J., Feng, J.H., D'Arcy, J.: Developing usable CAPTCHAs for blind users. In: Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility, pp. 245–246. ACM, New York (2007)
14. Ross, S.A., Halderman, J.A., Finkelstein, A.: Sketcha: a CAPTCHA based on line drawings of 3D models. In: Proceedings of the 19th International Conference on World Wide Web, pp. 821–830. ACM, New York (2010)
15. von Ahn, L., Maurer, B., McMillen, C., Abraham, D., Blum, M.: reCAPTCHA: Human-Based Character Recognition via Web Security Measures. Science 321(5895), 1465 (2008)
16. Bursztein, E., Bethard, S., Fabry, C., Mitchell, J.C., Jurafsky, D.: How good are humans at solving CAPTCHAs? A large scale evaluation. In: Proceedings of the 2010 IEEE Symposium on Security and Privacy (SP 2010), pp. 399–413. IEEE Computer Society, Washington, DC (2010)
17. Banday, M.T., Shah, N.A.: Challenges of CAPTCHA in the accessibility of Indian regional websites. In: Proceedings of the Fourth Annual ACM Bangalore Conference (COMPUTE 2011), pp. 31:1–31:4. ACM, New York (2011)
18. Lee, Y.L., Hsu, C.H.: Usability study of text-based CAPTCHAs. Displays 32, 81–86
19. Fidas, C., Voyiatzis, A.G., Avouris, N.M.: On the necessity of user-friendly CAPTCHA. In: Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems (CHI 2011), pp. 2623–2626 (2011)
20. Belk, M., Fidas, C.A., Germanakos, P., Samaras, G.: Do Cognitive Styles of User Affect Preference and Performance related to CAPTCHA? In: Proceedings of the 2012 Annual Conference on Human Factors in Computing Systems (CHI 2012), pp. 1487–1492 (2012)
21. Khalil, A., Abdallah, S., Ahmed, S., Hajjdiab, H.: Script Familiarity and Its Effect on CAPTCHA Usability: An Experiment with Arab Participants. International Journal of Web Portals 4(2), 74–87 (2012)
22. Wei, T.-E., Jeng, A.B., Lee, H.-M.: GeoCAPTCHA - A novel personalized CAPTCHA using geographic concept to defend against 3rd Party Human Attack. In: 2012 IEEE 31st International Performance Computing and Communications Conference (IPCCC), pp. 392–399 (2012)

Designing for the Functionality South African Internet Banking Websites Should Provide to Address the Needs of Generation-Y Users

Sebabatso Mtimkulu, Judy van Biljon, and Tobias van Dyk

School of Computing, University of South Africa, South Africa
sebabatsoa@aquaonline.com, {vbiljja,vdyktj}@unisa.ac.za

Abstract. Despite the widespread adoption of Internet banking there are no validated guidelines on the functionality the younger, techno-savvy Generation-Y customer segment (18-35 year age bracket) expect from Internet banking websites. This research investigated the functionality the Generation-Y customer segment require from South-African Internet banking websites. The User Centred Design (UCD) philosophy with a mixed method research design was utilised. Generation-Y technological characteristics and preferences abstracted from the literature were aligned with functionality trends of future Internet banking websites to formulate an initial list of Generation-Y aligned Internet banking functionality guidelines. These were evaluated during interviews with representative Generation-Y customers and also used in the heuristic evaluation of the Internet banking platforms of five South African banks. The findings were integrated towards synthesizing functionality guidelines. A visual representation of these functionality guidelines was constructed as a wireframe prototype for evaluation by Generation-Y users. The main contribution of the study is the validated list of Internet banking functionality guidelines for Generation-Y banking customers.

Keywords: Functionality, Internet banking, Generation-Y, User Experience.

1 Introduction

Internet banking is the second most used service on the Internet in South Africa, the most-used service being email [1]. In the face of expanding ecosystems of legacy websites, mobile sites and applications, banking institutions need guidelines to help best align their users' experiences with their expectations. A number of studies have been performed around the expectations users have of the Internet banking platform, but none of them have explicitly outlined functionality guidelines for the strategically important and strongly differentiated Generation-Y customer segment. Unlike any other predecessor generation, Generation-Y members are the first generation to be brought up with ubiquitous exposure to digital technologies, and because of this, have adopted technology as a primary tool for communication, education, as well as information gathering and sharing [2]; all qualities that shape the perceptions they have of

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 366–383, 2013.

websites [3]. Given these unique characteristics, providing relevant online experiences for them is important, as it is likely to translate into profitable future investments for organisations [4].

Research evaluating functionality on the Internet banking websites of six of the largest banks in the United States (Bank of America, Chase, Citibank, PNC Bank, U.S. Bank, and Wells Fargo) [5], uses a Website User Experience and Functionality Benchmark methodology to evaluate how the different banks measure against an already established set of Internet banking functionality guidelines. These guidelines outline conventional functionality currently existing on Internet banking platforms, such as paying beneficiaries, inter-account transfers, etc., but do not explore the next level of functionality innovation this platform can offer to banking customers.

In his research, Ravendram [6] identifies a single advanced functionality component of Internet banking in the Australian context, namely *customisation*. Herein, he explains that customisation is an imperative functionality dimension for the Internet banking platform, particularly among the younger generation, and continues to stipulate relevant technology that may assist to bring this dimension to life. Although customisation is identified as a vital functionality need, the paper does not identify additional functionality dimensions that may satisfy this younger segment of banking clients; nor does the paper reveal the characteristics of these banking clients that result in their inclination to customisation.

A study by Green and van Belle [7] investigated expectations of Internet banking in South Africa, and found that customers were satisfied with the basic Internet banking experience, but not with cost issues, speed, lack of integration with other banking channels and more advanced functionality. Besides identifying the need for advanced functionality, they do not expand on what this functionality could be.

The possible mismatch between the expectations of Generation-Y banking customers, and the functionality currently offered by South African Internet banking websites provide the rationale for this study. The research is guided by the following question: What functionality should South African Internet banking websites provide to address the needs of Generation-Y users?

Functionality, together with branding, usability and content has been identified as one of the key contributors to a positive user experience [8]. Therefore we argue that this study on Internet banking functionality contributes to the debate on human-computer user experience and especially designing for diversity.

Supporting documentation is available online to provide additional information that goes beyond the scope of this paper. Where relevant, these documents (referenced in the text as [23-25, 33, 36-37]), are cited for providing additional information related to this study. Section 2 describes the research design, while Section 3 sets the theoretical foundation and outlines the Generation-Y technological characteristics. Section 4 provides an overview of the literature leading to the formulation of the initial list of Internet banking functionality guidelines, while section 5 depicts how these functionality guidelines were validated. Section 6 discusses these findings, and leads to Section 7, which concludes the paper.

2 Research Design Overview

There are three types of research methods, namely qualitative, quantitative and mixed methods [9]. This study utilised mixed method research, with elements of both quantitative and qualitative approaches present, as demonstrated below in the next two paragraphs and in Fig. 1.

A literature review was conducted to identify the characteristics that influence the online preferences of Generation-Y users. A second literature review was conducted on the forecasted functionality of online financial services like Internet banking, to identify specific future functionality trends. These trends were then mapped to the Generation-Y characteristics previously identified to synthesise an initial list of innovative Internet banking functionality guidelines.

The philosophy of User Centred Design (UCD), advocates the active involvement of users for a clear understanding of user and task requirements [10]. Hence, a questionnaire was formulated and used in interviews with representative Generation-Y users, to validate the initial list of functionality guidelines gathered from the literature, and to identify new, desired functionality areas. The initial guidelines were then consolidated with the additional user input from the interviews, to formulate a list of functionality heuristics that were used to inspect the Internet banking websites of five banks in South Africa (Standard Bank, ABSA, FNB, Nedbank and Capitec).

This inspection, investigated whether the functionality currently available on the banks' Internet banking websites met the gathered Generation-Y functionality guidelines while seeking to uncover existing functionality deemed as pioneering, and worthy of being part of the Internet banking platform of the future.

A set of updated Generation-Y Internet banking functionality guidelines was synthesised from the reports and visually represented in the form of a semi-functional, HTML wireframe prototype. Generation-Y Internet banking users evaluated the prototype to validate the guidelines. The research design flow is depicted in Fig 1, which shows the three main data capturing methods, used (interviews, heuristic evaluation and prototyping), and which are further discussed in section 5.

3 Identifying Generation-Y Characteristics

"Today's multichannel customers demand better experiences than they get from institutions that design underperforming, one-off touch points … Institutions need a plan that will help them align their investments with their customers' most pressing needs. How can they do this? They can accomplish this by taking a user-centred approach to understanding the needs and behaviours of their customers and filling in the experience gaps" [11:1]. This quote emphasises the importance of understanding how diverse user groups' expectations and behaviour influences their perceptions of technology and consequently online banking platforms. Generation-Y technological characteristics, supporting specific online inclinations they have are discussed in Section 3.1.

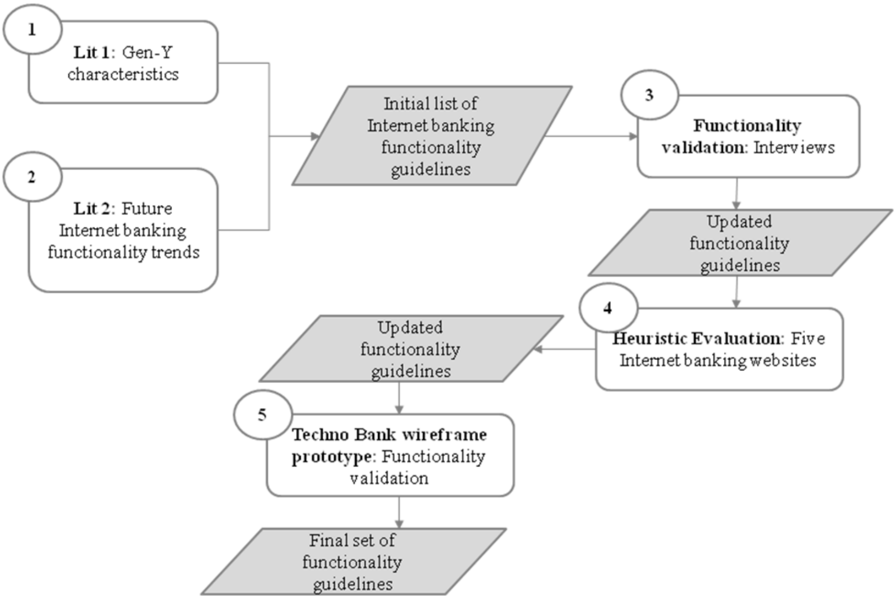


Fig. 1. Phases leading to the synthesis of the final Internet banking functionality guidelines

3.1 Technological Characteristics of Generation-Y Users

Generation-Y users are likely to own multiple digital devices as they desire continuous connectedness [12]. From this, general mannerisms such as "SMS language" have emerged, resulting in recommendations that online platforms speak to them in a tone they can relate to, always seeking to shorten and simplify instructions, and where possible use visual cues instead of long text [3]. They are mostly influenced by peers and family, are continuously looking for online human connections and exchange of real experiences that makes them feel a sense of community, respect and acceptance [13]. Tools of expression that enable them to communicate with others are therefore recommended on any online platform aiming to cater for this user base.

These users expect instant gratification, and expect online processes to execute promptly. Therefore, websites have to be designed with immediacy in mind, exposing the value of key pages and regularly updating content, whilst providing constant feedback to these users [14]. These users are more than mere consumers of online content, but are also active creators, continuously updating online blogs and sharing generic life events. Because of their diverse and expressive nature, they expect to be provided with high levels of personalisation and customisation where they are able to change interactions, and even products and services to reflect their individual personalities [3]. These discussed technological characteristics can be summarised into three main categories with supporting sub-categories as shown in Table 1.

Table 1. Technological characteristics of Generation-Y users [3]

Category 1: Technologically fluid and highly networked	
Sub-characteristic	Explanation
Continually connected	These are owners and users of multiple technology gadgets like portable MP3 players, laptops, tablet devices, smartphones, etc. and never want to be seen as "out of touch" [12].
Category 2: Emotionally looking for connections	
Sub-characteristic	Explanation
Influenced by peers	This generation highly relies on recommendations from friends and family, hence the high prevalence of use on social networking platforms such as Facebook. Therefore, online platforms should seek to provide self-expression tools that will allow these users to communicate with others.
Enjoy humour with an odd slant	Most of these users spend time actively seeking online experiences that are both humorous and entertaining.
Speak their own language	Generation-Y users are influenced by communication technologies like chat, text and instant messaging. Because of this, this user base has developed their own language ("SMS language") comprising of acronyms, emoticons, modified spelling, and slang. It is recommend that online platforms speak in a tone that authentically addresses these users, as well as keep instructions simple, using shorter sentences, and where possible substituting long text with images and other interactive visuals [14].
Category 3: Unpredictable and creative	
Sub-characteristic	Explanation
Skim content very quickly, and are easily bored	This generation is used to instant technology, and therefore quickly scans through information and rapidly absorb content, without dwelling on or reading any text-heavy pages. Online platforms designed for Generation users should therefore expose immediate value on key pages of the website, as well as update and refresh content, whilst regularly providing them with feedback [14].
Expressive	Most of these users are active content creators, continuously updating blogs, uploading videos, sharing pictures, etc. They are also expressive and respond to online experiences that provide them with the freedom to personalise and customise interactions and products and services in a manner they believe befit their individual personalities.

4 Identifying the Functionality Trends and Guidelines

4.1 Internet Banking Functionality Trends

Website functionality is an important part of the way an organisation does business by using a computer [15]. The use of digital technology provides access to those extensions of an organisation's services that can be accessed at the customer's convenience. Customers now prefer to pay bills, apply for services, initiate bill disputes and more, all from the convenience of their home or office [16]. This is made possible by the functionality offered on various online self-service platforms.

As noted, functionality is one of the four key contributors to a positive user experience (together with branding, usability and content) [8]. Functionality needs to cater for all technical and task supporting processes and applications that entail the delivery of the website's interactive services according to the unique needs of the user on the site.

The analysis revealed seven distinct functionality trend categories as pioneering and definitive of future digital financial services such as Internet banking. These are personal financial management (PFM) [17], multi-device banking [18], personalisation [19], process automation [19], content presentation [20], human touch [21], and social banking [22]. The analysis also revealed early-adopter banking institutions around the world already implementing some of these functionality guidelines [23]. These include Mybarclaycard (UK), BBVA (Spain), Jyske Bank's (Denmark), Commonwealth Bank (Australia), 22Seven, Standard Bank, Nedbank and FNB (South Africa), as well as Hapoalim Bank (Israel). Explanations of what each of the seven functionality categories represent are given below.

- *Personal financial management (PFM):* This alludes to transactional banking users being able to manage their money by using money aggregation technology that enables them to have a consolidated view of their financial standing, usually across a number of financial services providers. Where relevant, users are provided with personalised budgeting and money management hints and tips.
- *Multi-device banking:* Due to the prevalence of portable mobile devices such as smartphones and tablets, banking organisations have to ensure that their online platforms can be viewed comfortably, regardless of the device being used.
- *Personalisation:* Future digital financial services will award users the opportunity to determine their own website settings in a manner that befits their personal preferences (e.g. change of interface theme, navigation rearrangement, etc.). This platform will also learn about the user's financial profile, and recommend relevant products, website content and functionality.
- *Content presentation:* It is no longer enough to have the right content; but content needs to be presented in more creative ways (e.g. video) that makes it easy to read and comprehend, resulting in users quickly absorbing key details.
- *Process automation:* This category addresses banking functionality that will enable users to perform certain transactions, from start to finish, without having to access a physical bank branch. These transactions include increasing of certain account limits, product applications, etc.

- *Human touch:* Banking technology should be balanced with human interaction, by allowing users to access human assistance from the bank whenever the need arises.
- *Social banking:* Digital financial services of the future will demand integration across several processes, systems, applications, and channels; with social media being part of this equation.

4.2 Initial Generation-Y Functionality Guidelines

Having discussed the Generation-Y characteristics (Section 3.1) and the projected functionality trends for banking websites (Section 4.1), we now consider the desired user experience on future digital financial services. This can be summarised in terms of the S.U.P.E.R. acronym [19] where the meaning of the letter can be explained as follows:

- *Simple:* It will be much easier for customers to achieve goals and tasks on digital financial websites, if products and services information are presented in a manner that is easy to comprehend.
- *Ubiquitous:* Customers will seamlessly interact with their financial services provider through an increasing number of touch points like mobile devices and social networks. There will be continuity and consistency across all these platforms, making users comfortable and assured that their needs are met regardless of the device and platform they choose to transact on.
- *Personal:* The entire online experience will be relevant to unique customer needs, and will not use a one-size fits all approach when it comes to the handling of customers' financial data.
- *Empowering:* Customers will be able to take action by themselves, as they will be provided with an aggregated view of their finances, from multiple sources. This aggregation will put them in control of their financial lives, as they will have a single and solid point of reference for all their financial information.
- *Reassuring:* As human beings still remain the best sales and service channel for many high-value interactions, future digital financial services will provide human help whenever the user feels the need.

With acquired knowledge of Generation-Y characteristics (Table 1), future functionality trend categories (Section 4.1), and an overview of the expected user experience on digital financial services (S.U.P.E.R [19]), it became possible to draw similarities between these different sources of information from the literature, and assess whether the functionality trends and the projected user experience aligned with the gathered traits of Generation-Y users.

Table 2 depicts the triangulation between these three aspects, mapping the functionality trends and user experience elements to the most relevant Generation-Y technological characteristics.

The projected user experience, and functionality categories align in many ways with the technological traits of Generation-Y users, accentuating that the type of functionality trends gathered from the literature could potentially be appealing to this users base. This was validated during interviews with Generation-Y users (Section 5.1).

Due to its functionality encompassing nature, the S.U.P.E.R acronym was used as main categories, in which the seven functionality trend categories (Section 4.1) were grouped. From these trend categories, an initial list of thirty specific functionality guidelines was synthesized for use during the user validation (interview) phase of the study [24].

Table 2. Triangulation of functionality trend categories, S.U.P.E.R and Generation-Y characteristics

Projected user experience on digital financial services	Functionality trend	Generation-Y technological characteristic
Simple	Process automation and content presentation	Skim content very quickly, and are easily bored
Ubiquitous	Multi device banking and social banking	Continually connected and influenced by peers
Personal	Personalisation	Expressive and creative, speak their own language, and enjoy humour with an odd slant
Empowering	Personal financial management	Expressive and creative
Reassuring	Human touch	Influenced by peers

5 Validating the Functionality Guidelines

To validate the characteristics and functionality guidelines, interviews, heuristic evaluation and prototype evaluation were used.

5.1 Interviews

Interviews were conducted with representative Generation-Y users. Participants were sampled from AquaOnline (Pty) Ltd., a full service digital and direct marketing agency. The average age of employees in this organisation is 31 years. Due to the nature of the business, employees are required to have a wide knowledge of and exposure to digital technology in general, positioning the company to provide a representative sample of the Generation-Y population at large. Twelve Generation-Y participants were sourced, with the questionnaire utilised during the interviews [25] designed to:

- Validate the demographic details of the participants.
- Gather perceptions they currently have of their respective Internet banking platforms.
- Validate the literature gathered list of thirty Internet banking innovative functionality guidelines.
- Ascertain whether participants could identify any additional functionality innovation not yet uncovered by the thirty guidelines [24] presented to them.

A rating scale is the most efficient manner of capturing self-reported data [10]. In order for each participant to demonstrate their level of agreement or disagreement with the functionality guidelines presented to them [24], a 5-point Likert scale was used. Participants were requested to rate the different guidelines, and where relevant, provide substantiation for their rating. The questionnaire design allowed for both qualitative (participants' substantiation on functionality rating, and additional functionality recommendations from them) and quantitative data (participants' functionality acceptance rating) capturing, thus providing a holistic understanding of the importance participants were placing on the different functionality guidelines presented to them.

The interviews exposed the perception participants currently have of their Internet banking platforms, with their input revealing that most are happy with the basic transactional capabilities this platform offers (i.e. paying beneficiaries, buying airtime, inter-account transfers, etc.). Even though this is the case, they acknowledged that the functionality on this platform could be improved, by providing them with the ability to better manage their financial lives. This was further supported by the participants' views that they believe this platform should be playing a role of a "financial adviser", which it currently does not fulfil.

When asked to validate the functionality guidelines presented within the various categories (i.e. simple, ubiquitous, personal, empowerment; reassuring), it was revealed that Generation-Y users are looking for full process automation, where they would be able to perform actions that currently require a branch visit (e.g. increasing transactional limits) online. They are wary of security complications the introduction of functions like electronic signatures may introduce, and strongly communicated the need for additional layers of security like One Time Passwords (OTPs) should such functionality be incorporated. This input yielded an acceptance rating of 4.2 (agree with proposed functionality), for functionality listed within the Simple category.

Participants agreed with the idea of being notified about other channels of banking (e.g. via a mobile app) however, being able to undertake their banking on social networking websites like the notion of social banking suggested was unsettling to them. The functionality listed within the Ubiquitous category therefore yielded an acceptance rating of 3.2 (Neither agree, nor disagree with proposed functionality).

Even though they supported the idea of customising the platform according to their needs (e.g. change of interface colour, navigation rearrangement, etc.), participants communicated that what users who are permitted to modify on the platform interface should be closely monitored. They felt it could interfere with the overall branding of the organisation; a component of the interface they believe is key in depicting the stature and credibility of the banking institution. An inclination towards personalisation, rather than customisation was observed, where they preferred the bank to make targeted promotions based on products they do not currently have, rather than allowing them to change the background colour of their transacting interface. The Personal category therefore returned a 3.8 acceptance rating (agree with proposed functionality).

Participants were keen to see the transformation of the Internet banking platform into a money aggregation site, where they would be able to have a better view of, and manage their finances more efficiently. They want to be able to set and monitor savings targets, put spending alerts into place, budget online, track their spending on certain categories like groceries, entertainment, fuel, airtime etc. The empowerment category averaged at a 4.6 acceptance rating (strongly agree with proposed functionality). Participants also supported the idea of having direct human contact on the platform, with the click-to-chat and request a call back functions being the most preferred. The Reassuring category yielded a 3.8 acceptance rating (agree with proposed functionality).

When asked to identify additional functionality not listed in the guidelines, Generation-Y users communicated that they want to be able to access the Internet banking platform more easily, by being able to select their own username and passwords; and not have to memorise lengthy card and profile numbers. They are looking at this platform to start offering more than just banking solutions, but start addressing other aspects of their lives (e.g. what a healthy combination of financial products is, and where the users are falling short). They are expecting this platform to offer financial education not necessarily related to their existing product set (e.g. share trading tutorials, investment tips, etc.), as well as make product offers not necessarily related to banking (e.g. discounted holidays, sales on certain makes of mobile devices, etc.).

To them, the Internet banking platform is an extension of other electronic based ways of transacting, thus they want to easily extend their transactions on a mobile device of their choice (e.g. on a smartphone or tablet), while on the move, without the experience being tainted. Overall, Generation-Y participants require their online transacting platform to be a one-stop-shop that has their overall financial well-being at the core of the offering, while slowly starting to creatively integrate other parts of their lives.

5.2 Heuristic Evaluation

The functionality guidelines validated and further recommended by participants during the interviews were then used to formulate a list of heuristics that was used to inspect the Internet banking websites of Standard Bank [26], FNB [27], ABSA [28], Nedbank [29] and Capitec Bank [30]. Although a commonly known set of Nielsen's usability principles [31] is usually used during a heuristic evaluation, Pinelle, Wong and Stach [32] advocate that a heuristic evaluation is flexible and versatile enough to be adapted to specialised domains [32].

Five expert evaluators, all of them experienced user experience professionals accustomed to the heuristic evaluation process, took part in the evaluation [33]. Each evaluator compiled a heuristic report on their findings, stipulating how they felt each of the Internet banking websites faired against the functionality guidelines gathered thus far. The evaluation also identified functionality on current Internet banking platforms the evaluators deemed worthy of being labelled "innovative".

For Standard Bank, FNB and Capitec Bank, transactional capabilities are still the main focus of the platforms, with ABSA and Nedbank breaking the norm by beginning to position their Internet banking websites as interactive money management platforms. These two banks have started to introduce, to a certain extent, personal financial management capabilities, and offer users generic financial education that is aimed at helping them understand various financial topics. The account aggregation ability (i.e. a consolidation across multiple accounts), is still manual on the ABSA site, and on a platform outside Internet banking for Nedbank. However, the general consensus from evaluators was that these two banks did at least plant a seed of advanced financial consciousness in their users' mind, and allows the websites to be positioned as more than a platform where basic transactions happen, but rather "partners in their financial well-being".

FNB is making headway on the cross selling of products on the site. Evaluators identified an additional functionality type they deemed original on the websites of this bank, as well as Nedbank. The functionality enables users to rename accounts that appear on their transacting homepage, effectively making products they have with the bank easily memorable and identifiable. None of the five banks evaluated satisfactorily portrayed all the recommended functionality guidelines in a manner that could comprehensively cater for the identified needs of Generation-Y users.

Although Nedbank and ABSA show potential, their platforms are not yet a seamless representation of all the functionality guidelines gathered in this study. All the evaluators emphasised the value and importance of this study for banking institutions wishing to target this particular customer segment.

5.3 Functionality Validation with Wireframes

Wireframes are schematic presentations that define a webpage's content and functionality structure in order to portray the page concept before it is designed and developed [34]. Based on the functionality guidelines gathered and updated throughout the different phases of the study (literature, interviews and heuristic evaluation) a semi-functional, HTML wireframe prototype for a fictitious bank ("Techno Bank") was created [35]. Fig. 2 depicts a screenshot of the money management section of the prototype; an entry point to the empowering category functionality guidelines. For instance, in this section of the prototype, a user is able to see a consolidated view of all their financial accounts, and is also able to track their spending progress over a certain amount of time; all functionality validated and deemed important by Generation-y participants during the interviews.

Functionality representing the other four requirement groups (i.e. simple, ubiquitous, personal; reassuring) is also contained in the prototype. Users are able to, amongst other things, view bank products recommend to them based on their unique financial profile, access financial education in order to improve their financial standing, download banking applications (apps) for other devices like their tablets and smartphones, and initiate an instant chat session with their respective banker; all functionality they supported during the validation phases of the study. Ten representative Generation-Y users were requested to access this prototype, and validate whether they would be satisfied with the functionality proposed, on their actual Internet banking platforms. Participants were given tasks [36], representative of all the gathered functionality guidelines to ensure they were familiar with the different functionality guidelines, thus enabling them to provide informed functionality validation feedback. Similarly to the functionality validation process during the interviews, participants herein also demonstrated their level of agreement or disagreement with the prototype functionality by utilising, a 5-point Likert scale.

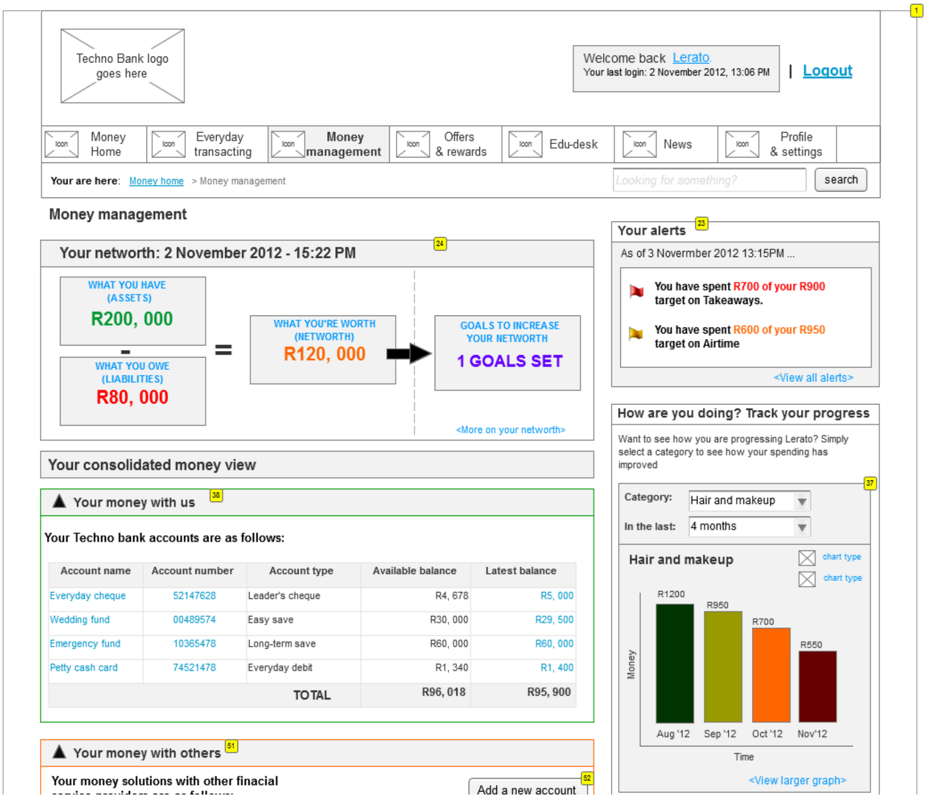


Fig. 2. Example of the Techno Bank *empowerment* section of the prototype

Overall, participants reached a consensus that functionality projected in the "Techno Bank" prototype is the kind they would like to have on their Internet banking platform of the future. For each functionality category, a functionality acceptance rating was captured, with participants further elaborating on certain aspects of the proposed functionality. Their input for each functionality category is summarised below.

- *Simple* (Acceptance rating 4.6: Strongly agree with proposed functionality): Participants agreed that simplifying transactions and the consumption of product related information were important functions to have on an Internet banking site. They communicated that having direct access to calls-to-action such as ''apply now'' was very important for them on the site. They expect to action their needs immediately when a tool or product presented is of interest to them. They also supported the idea of the bank reusing personal information (i.e. their full name, identity number, income and address details) it already has about them to help them complete certain online processes (e.g. applying for a new product) quicker. Other media of communication such as video were also supported, with users communicating that this is a good alternative for those who do not wish to read lengthy product details. When asked to validate the ability for users to action a transaction by using an electronic signature, users were concerned and agreed only on condition that an extra layer of security (such as a One Time Password), be provided. Although this functionality innovation would further automate transactions, users surprisingly expressed the view that they would still prefer going into a bank branch to render a physical signature. Issues of security and trust remain a concern to them despite their early exposure to technology.

- *Ubiquitous* (Acceptance rating 5: Strongly agree with proposed functionality): All users agreed that being notified of other ways of banking, as well as being able to instantly download any new apps on the site was a useful feature to have. They also communicated that they would like the banks' apps to be linked to their mobile app store, where they would be able to be notified of any new app developments the bank is introducing. All participants agreed on excluding social media integration from the platform. They contended that this should not be incorporated at all, even in more subtle ways like the sharing of an interesting educational article posted by the bank in the education section of the prototype.
- *Personal* (Acceptance rating 4.8: Strongly agree with proposed functionality): Users noted that any promotion the bank extends to them should add value to their financial lives. They spoke strongly against what they termed "irresponsible targeting", where the bank pushes promotion messages without taking a closer look at what their actual financial needs really are. Generally, they supported the idea of being able to customise the site by rearranging certain interface elements, as well as renaming the different bank accounts the user has. Changing the background of the site's interface to a different colour or theme was, however, met with scepticism, with most of the users expressing that they felt this is one aspect of the bank that depicts its credibility, and therefore allowing users to have too much leeway with this, would mean the bank letting go of a sense of trustworthiness usually presented by its brand identity (i.e. certain logo, colours, etc.).
- *Empowering* (Acceptance rating 5: Strongly agree with proposed functionality): All users agreed that having a consolidated view of their finances would help make them conscious of how they spend, and act as a catalyst to help them make better financial decisions. They supported the functionality that allows them to track their spending within a particular category (e.g. hair and makeup, groceries, fuel, etc.). Here, they expressed they would like to compare multiple categories simultaneously. They also appreciated the setting up of an online budget, and herein expressed the need for this function to be synchronized to a mobile device of their choice, for them to conveniently access it when they are away from a laptop or desktop computer.
- *Reassuring* (Acceptance rating 5: Strongly agree with proposed functionality): Users appreciated the ability to chat directly with a bank representative. For this particular functionality, they expressed the need of having an upload function, where they would be able to attach documentation while the online chat with the bank representative was taking place. They supported the idea of being able to leave their details for the bank to contact them, as well as sending and receiving secure messages from the bank.

- *Additional functionality guidelines* communicated by users during the interview phase of the study (Acceptance rating 5: Strongly agree with proposed functionality): Users supported the introduction of more than just transactional content on the platform. They all agreed with the idea of being able to select their own login details, as they communicated this would be a more efficient option of entry. They were satisfied with the introduction of a rewards section on the prototype, where they would be able to use their loyalty points on discounted products not necessarily related to banking. They were accepting of the notification of technologies like Near Field Communication, or Geo-payments as it is commonly known, as they believe the Internet banking platform should be a central hub, giving them visibility of all other possible ways of banking; a function that is currently not available on the platform.

6 Discussion

Having considered traits supporting Generation-Y inclinations to certain online experiences (Section 3.1), an initial, literature-based list of thirty innovative functionality guidelines was formulated [24] (Section 4.2). The research process was designed so that the output from each phase informed the next. Therefore, the initial functionality guidelines [24] were refined and validated by the various phases (interviews, heuristic evaluation and prototype evaluation) to provide a well-developed and thoroughly scrutinized list of functionality guidelines for the Internet banking platform [37]. As the journey of functionality validation unfolded, key feedback patterns were observed, leading to a set of ten, high level functionality implementation guidelines supportive of the detailed functionality guidelines. When designing Internet banking functionality targeted at the Generation-Y user base, the designer should ideally:

1. *Provide a one-stop-shop of financial guidance:* Generation-Y users are looking for more than a transaction platform; they expect a financial partner that will assist in taking care of the holistic needs of their financial lives. The platform therefore needs to take the lead and provide them with a comprehensive view of their financial standing, while guiding and equipping them on how to better their financial situation.
2. *Target with a purpose:* The techno-savvy nature of this user base makes them conscious of marketing messages that are planted without adding specific value. Whatever is targeted or cross-sold to them should therefore demonstrate value, and a contribution towards a better financial standing.
3. *Pull in resources to educate:* Where relevant, provide external, third party educational resources not necessarily compiled by the bank. Participants explained that the bank doing this demonstrates the value the bank places on their customers' overall financial well-being.
4. *Provide cross-channel experiences:* Users are expecting to action a transaction on the Internet banking platform, and have this readily available on a mobile device of their choice. To them the experience should be seamless and continue from one

channel to another. They should therefore never feel like there's something one channel provides, that they cannot access and execute on another.

5. *Design with immediacy in mind:* These users are looking for quick, easy and convenient ways of consuming content and executing tasks on this platform. The use of video as an alternative to heavy text, or any other more convenient forms of consuming content should be utilised. The placement of "the next best action" should be well thought of, as these users want to action immediately, or as soon as a product or content type catches their eye.
6. *Automate wherever possible:* Reduce, as far as possible the need to access a physical bank branch. These users are expecting this platform to offer capabilities that simplify and make their banking lives convenient.
7. *Personalise and customise within limits:* These users want to be in full control of the journey on the site, also controlling the rendering of the transactional interface. When addressing the latter, allow them to modify certain aspects that will not hamper the overall brand identity (e.g. colours and logo) of the bank, as according to them, this preserves the sites credibility.
8. *Offer rewards for being loyal:* Show users the benefit of staying with the bank, by rewarding them for being part of the establishment. Award them the opportunity to be able to redeem these loyalty rewards on their Internet banking platform. Create partnerships that are not necessarily related to banking (e.g. for smartphone, holidays, restaurants, etc.), and offer these deals to users.
9. *Not slack on security and privacy:* As functionality recommendations were made on this platform, users became more and more concerned about the security, hence the requesting of an additional authentication layer for some of the proposed functions (e.g. incorporation of electronic signatures to complete a transaction). Therefore, a tighter implementation of security measures like the One Time Password (OTP) should be prioritised.
10. *Speak casually, yet authoritatively:* Adopt a tone that is friendly, yet formal when addressing this user base. They require a platform that is free of financial jargon, while being able to provide them with the financial guidance they expect from their bank.

7 Conclusion

This paper reported on an investigation into the functionality the Generation-Y customer segment require from South-African Internet banking websites. Previous research on Internet banking functionality investigated how the different banks measure against an already established set of Internet banking functionality guidelines [5, 6].

A South African study identified the lack of advanced functionality on Internet banking websites as a problem [7] but none of these addressed the problem of providing Internet banking functionality for the Generation-Y customer segment. Therefore the validated guidelines for understanding Generation-Y users' functionality as presented in this study, contribute to the holistic understanding of this user groups' needs, behaviours and expectations.

Furthermore, it provides practical, user-centred functionality guidelines that can result in the immediate alignment of Internet banking functionality to Generation-Y users' needs; a novel approach to Internet banking research, not yet ventured into by any of the banks investigated. Deeper insight into the unique characteristics of this segment also means that they can be utilised to gather an understanding of these users for any other web interface effort not necessarily related to banking. The literature found on Generation-Y needs did not differentiate based on biographical characteristics such as socio-economic status and education within the age group. That is a limitation that needs to be addressed in future research on Generation-Y characteristics.

Although particularly focused on the South African market, functionality trends and innovation initially gathered in the literature are from both local and global best practice (as outlined in Section 4.1), and can therefore be utilised as a benchmark for online banking platforms outside the South African context. Furthermore, although South African Generation-Y users were interviewed, and were requested to validate the proposed functionality, members of this generation have been equally impacted by globalisation and international influences, and therefore the same set of attributes can be used to describe them worldwide [38].

This research has revealed a dire need for broader, structured research into Internet banking functionality needs, and this is reflected in the paucity of research papers on this topic. Further research is required to validate the guidelines with a larger sample of the Generation-Y segment, and to investigate additional innovative functionality not covered by this study.

References

1. Muller, R.: What South Africans do online, http://mybroadband.co.za/news/internet/26873-what-south-africans-do-online.html
2. Smith, R., Cha, V.: Understanding Generation Y and their perception of e-Government, http://www.egl.sg/downloads/Gen_Y_and_their_Perception_of_e-Government.pdf
3. Temkin, B.D., Popoff-Walker, R.: The Gen Y design guide. Forrester Research Inc., Massachusetts (2007)
4. Oracle Financial Services: Are banks ready for the next generation customer? http://www.oracle.com/us/industries/financial-services/gen-y-survey-report-165297.pdf
5. Wannemacher, P.: 2011 US Bank Secure Website Rankings. Forrester Research Inc., Massachusetts (2011)
6. Ravendran, R.: Website Customization: Exploring a Tag-Based Approach in the Australian Banking Context. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) INTERACT 2011, Part IV. LNCS, vol. 6949, pp. 434–437. Springer, Heidelberg (2011)
7. Green, S., van Belle, J.: Customer expectations of Internet banking in South Africa, http://www.commerce.uct.ac.za/informationsystems/staff/personalpages/jvbelle/pubs/f-VanBelleJeanPaul2.pdf
8. Rubinnof, R.: How to quantify the User Experience, http://blogs.sitepoint.com/quantify-user-experience/

9. Creswell, J.: Research Design – Qualitative, Quantitative and Mixed Methods Approaches. Sage Publications, Los Angeles (2009)
10. Tullis, T., Albert, B.: Measuring the User Experience – Collecting, Analyzing and Presenting Usability Metrics. Elsevier, Massachusetts (2008)
11. Rogowski, R., Manning, H., Stone, A.: Improve your digital customer experience. Executive overview: The digital customer experience improvement playbook. Forrester Research Inc., Massachusetts (2011)
12. Patterson, G.: Marketing to Gen X and Gen Y, http://www.yr.com/sites/vmldev.com/files/Gen%20X%20%26%20Y%5B3%5D.pdf
13. McCrindle.: Understanding Generation Y, http://ec-web.elthamcollege.vic.edu.au/principal/pdf/Understanding%20Generation%20Y.pdf
14. Temkin, B.D., Mclnnes, A., Zinser, R.: Engage Gen Y online with immediacy. Forrester Research Inc., Massachusetts (2008)
15. Boiko, B.: Functionality is content too, http://206.253.219.101/biblev1/Whitepapers/Boiko_Whitepaper_WPCh4.pdf
16. Goldstuck, A.: Customer self-service strategies in South Africa 2010. World Wide Worx (PTY) Ltd & Consology (PTY) Ltd., South Africa (2010)
17. Niemeyer, V.: Case Study: How Barclaycard helps customers manage their spending. Forrester Research Inc., Massachusetts (2011)
18. Kissoyan, V.: To App or not to App? Why responsive design is key to your mobile strategy, http://www.lokion.com/lokion/to-app-or-not-to-app-why-responsive-design-is-key-to-your-mobile-strategy/
19. Hesse, A.: Next-generation digital financial services - Make it simple, ubiquitous, personal, empowering, and reassuring. Forrester Research Inc., Massachusetts (2011)
20. Ensor, B., Poltermann, S.: Using video to drive online financial services sales. Forrester Research Inc., Massachusetts (2011)
21. Montez, T.: Case study: Hapoalim Injects a human touch into digital banking. Forrester Research Inc., Massachusetts (2012)
22. Naidu, P.: Social media and its application within the banking industry. MSc Thesis, University of Pretoria (2010)
23. Examples of functionality innovation demonstrated by early-adopter banks, http://osprey.unisa.ac.za/TechnicalReports/Eg_func_innov_banks.pdf
24. Initial list of thirty Internet banking functionality guidelines, http://osprey.unisa.ac.za/TechnicalReports/init_30_func_guide.pdf
25. Example of participant interview questionnaire, http://osprey.unisa.ac.za/TechnicalReports/eg_interv_question.pdf
26. Standard Bank of South Africa, http://standardbank.co.za
27. FNB (First National Bank), http://fnb.co.za
28. ABSA, http://absa.co.za
29. Nedbank, http://nedbank.co.za
30. Capitec bank, http://capitecbank.co.za
31. 10 usability heuristics for user interface design, http://www.nngroup.com/articles/ten-usability-heuristics/
32. Pinelle, D., Wong, N., Stach, T.: Heuristic evaluation for games: Usability principles for video game design. In: ACM CHI Conference on Human Factors in Computing Systems, Italy, pp. 1453–1462 (2008)
33. Heuristic evaluation guidelines, http://osprey.unisa.ac.za/TechnicalReports/heu_eval_guide.pdf

34. Glen: Website wireframes: Samples and examples, http://www.fatpurple.com/2010/03/04/web-site-wireframes-samples-examples/
35. Techno Bank semi-functional HTML prototype, http://share.axure.com/BRQ8Y6/
36. Prototype functionality evaluation tasks, http://osprey.unisa.ac.za/TechnicalReports/func_valid_tasks.pdf
37. Final list of Internet banking functionality guidelines for Generation-Y users, http://osprey.unisa.ac.za/TechnicalReports/final_func_guide.pdf
38. Smith, C.: Black to the future - South Africa's Gen-Y, http://www.tomorrowtoday.co.za/2010/03/02/"black-to-the-future"-south-africa's-gen-y/

“Nobody Other Than Me Knows What I Want”: Customizing a Sports Watch

Piia Nurkka

Tampere University of Technology, Korkeakoulunkatu 1, 33101 Tampere, Finland
piia.nurkka@tut.fi

Abstract. In order for companies to effectively use customization as a design strategy, there should be understanding on what users would like to customize and why. This study explores the use of customization features of sports watch in order to assess the extent of customization, and to identify reasons for customization in this context. Survey data from 100 users of a sports watch were analyzed to understand how they use the different customization features: general preferences, functionality and appearance. The findings show that although the users vary in the use of customization, they state similar reasons for customization: control, ease-of-use, increased effectiveness, and better fit to personal preferences. The motivation to customize in this context is for the most part related to autonomy: to the sense of control the user has by having the tool to adapt the product according to own preferences, wants and needs.

Keywords: customization, sports watch, satisfaction, user experience.

1 Introduction

Customization features of technology are designed to provide users with control over appearance and functioning, thus increasing its personal relevance to an individual [1]. The ability to customize is a means to establish a closer connection between users and products thereby enabling them to determine appropriate product characteristics for themselves, and to meet their functional and hedonistic needs [2].

Customization as a product feature is attractive with sensible promises of benefits to customers and companies alike. Customized and more accurate offerings can provide customers with superior value [3], facilitate positive experiences, and increase satisfaction with the product [4]. Therefore, personalization may lead to positive brand associations and increase brand loyalty [1]. This can translate to increased cash flows and profitability for companies [5].

Research on customization of technological products has for the most part been confined to certain products, service contexts or points of views. For example, much research on customization of mobile and smart phones has been conducted (e.g. [6, 7, 8]). In addition, in a desk top context studies on customization can be found on PCs [9], software [10], and web site interfaces [11]. On the other hand, research has often focused on a certain type of customization, with dominance of appearance customization over functional. The different types of customization deliver different

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 384–402, 2013.

value for the user: appearance customization provides aesthetic value [12], whereas functional customization provides functional value improvements. Functional value, for instance as interaction efficiency [6], may lead to fulfillment of basic human needs, such as control and autonomy [11].

This study explores the use of customization features of a sports watch. The aims were to identify the extent to which customization features are used, the kind of customization features used, and the reasons for their use of customization. In addition, we wanted to understand if and how users differ in relation to customization, and if customization affects satisfaction with the product. The study gives insight into customization in a new context and the results can be used to inform design of customization that aims at enhancing user experience.

The paper is organized as follows: Section 2 starts with the definition of customization used in this research, and then reviews the related work on the topic. Section 3 describes the current study. Section 4 presents the results of the study. First, the extent of customization is reviewed generally, and secondly, by the types of customization separately. Last, the reasons for customization are reported. Finally, the results and limitations of the study are discussed in Section 5.

2 Related Work

2.1 Defining Customization

In HCI and related literature different terms on the concept of customization have been used. The concept means a process that changes the functionality, interface, information content, or distinctiveness of a system to increase its personal relevance to an individual [9]. However, the definition does not state who the actor in the process is: the process can be either user or system initiated. Sunikka & Bragge [3] have reviewed the literature and suggest a personalization framework in which the user involvement in the process is at the basis and is what makes the distinction between the terms. Customization should be used if the personalization process is user (or customer) initiated, whereas personalization should be used if the system initiates the process. In user-initiated customization, the user himself manipulates aspects of his device or system or configures content of a website based on his individual needs and wants (e.g. changes the ringtone or installs a new application on a smartphone). On the contrary, in the system initiated personalization the system/device adapts content using information about the user or his behavior captured by the system/device. The personalization is implicit as the user may not be aware of it, whereas in customization the user controls the changes. Following Sunikka & Bragge's [3] suggestion, the term customization is used in this paper. Customization is defined as user-initiated modifications to the functionality, content or interface (functionality-based customization) or appearance of the device (appearance-based customization) to increase its personal relevance to an individual. When referring to the related literature, the term customization is used instead of personalization when the question is about user-initiated process even though the term personalization may have been used in the original work.

2.2 Use of Customization Features

Research suggests customization brings many positive aspects. However, to what extent do users actually use the offered customization options? Häkkilä & Chatfield [6] examined the process of customization of mobile phones with 60 users during the first months of usage. The results illustrate active customization of the phone, with most customization occurring shortly after using the new phone for the first time. They found 11 out of the measured 17 features were customized, and nearly all had customized the most commonly customized features: the ringing tone, audio profiles and background image. The least customized features were generally more complicated to configure. Customization happened not only with fun or style features, but also with functional phone settings. A strong motivation for customization was to make the device look and feel "own" by changing the appearance of it to match the user's personal style and to reflect his interaction preferences. Interestingly, Ventä et al [7] report a contrary finding regarding the appearance customization of a mobile phone. In their interview study of 40 mobile phone users, the users stated that they bought a phone with looks they liked, and thus did not want to modify it afterwards.

Furthermore, Tossell et al. [8] studied the customization of a smartphone with 24 participants during the first two months of usage. The aim was to examine the relationship between personalization, device usage and usability. The results show a high range of customization, but not across every customization feature measured. For instance, every user acquired a case for their smart phone, but the number of applications installed differed. In general, the results show there are differences between users in the use of customization: not all users customize their smartphones and females and males customize their smartphones differently. A relation between customization extent and usability was found: those who customized their phones more tended to rate it as more usable. In addition, the users who customized more also used their device for greater periods of time on a broader range of applications.

In a study by Page et al [10] the typical customizations made on a word processor were identified. Most users had made changes to their general preferences settings when only few customized the visual appearance of the interface. However, almost all, ninety-two percent of the users, had customized their software in some way. Most of the customization was done to facilitate the participants' work practices, and consequently, users who used the software most did the most customization.

To sum up, users vary a lot on how many features they customize. Features relating to appearance customization are often among the most customized. However, customization is also performed to increase ease of use. Consequently, high level of customization may increase the time of use of the product.

2.3 Reasons and Motivations to Customize

Customization, by letting users modify their products and services, assumes users are active actors rather than passive consumers. The question why to customize has therefore attracted researchers. For example, Oulasvirta & Blom [1] applied modern theories of motivation to explain user behavior. According to them, analyzing motivations is necessary because they determine the personal relevance that an act of customization carries. However, users differ in how strongly they express different motivations for customization. Oulasvirta & Blom [1] claim there is no special need for customization; rather there are context-independent basic needs that are idiosyncratically manifested motivations related to the use of a product's features. Specifically, users are willing to customize when the product involves and nurtures their psychological needs of autonomy, competence, and relatedness, taps into and extends their interests and preferences, and makes it possible for a user to transform a uniform technology to personally useful and enjoyable tool that can be used to improve and enjoy life and work. Thus, even though there is no need as such for customization, appropriate design of customization features is crucial: "It is through customization features that technology can help the user to align her inner motivational resources with her actions." [1, p.13].

In addition, Marathe & Sundar's [11] study on customization of a web portal reveals that customization has deep psychological value when users take time to engage in it and make changes according to their preferences. Specifically, their data show that customization is associated with two psychologically meaningful gratifications: sense of identity and sense of control. Their results demonstrate how customization leads to two distinct patterns of effects: while the effect of customization on sense of control was fully mediated by sense of identity, its effects on sense of identity was only partially mediated by sense of control. This finding leads to the conclusion that "having control is all about being able to express one's identity, but feeling a sense of identity has implications beyond feeling sense of control" [11, p.787]. From an interface design perspective, customization features should therefore offer possibilities for self-representation to express identity. UI designers need to consider what gratification each customization feature will imbue in the user: by fulfilling higher-order needs like enhancing identity and control, designer can create an interface that is an end in itself rather than a means.

2.4 Dispositions to Customize and Effects of Customization

The theory of personalization of appearance (TPA), developed by Blom & Monk [3] and elaborated by Monk & Blom [13], indicates the dispositions to personalize and the following effects of personalization on user. The theory has been derived from studies on the personalization of web portals, mobile phones and computer desktops. The dispositions are categorized as the user-, system- or context-dependent and each of them has individual dispositions. For example, user-dependent dispositions are: the frequency of use of the system, ownership of the system, knowledge of personalization and usage time. The system-dependent dispositions relate to the ease and cost of personalization and to the effectiveness of personalization items, i.e. their ability to elicit emotions or improve ease of us. Presence of a disposition may increase the personalization behavior, however, the authors do not explain how many or which of the dispositions are crucial for the personalization to happen. Nevertheless, personalization brings about socio-emotional effects on user. Authors have identified effects that can be divided into Cognitive, Enduring Emotional, Transient Emotional and Control Effects.

3 Study on Customizing a Sports Watch

An online questionnaire was designed to provide information on the extent and the reasons to customize a sports watch (Figure 1.). In this section, first, the product and measured customization is explained, followed by the description of the online questionnaire, respondents and analysis of the data.

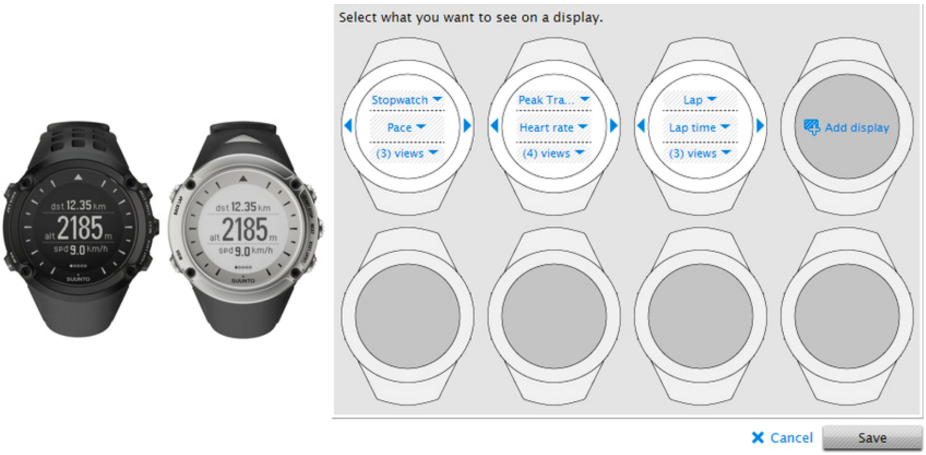


Fig. 1. One the left, the two different variants of the sports watch: the black with dark display and the silver with light display. On the right, a screen capture of the customization of running mode on the online training log

3.1 Customizing the Sports Watch

The product under investigation is a sports watch called Ambit produced by Suunto a Finnish developer and manufacturer of dive computers, heart rate monitors and outdoor sports instruments. The watch has an integrated GPS and the possibility to monitor heart rate. In addition, it has diverse set of specialized outdoor and training functions and is intended for (according to the marketing of the product) "the Outdoor Explorer" [14]. The product is a part of a cross-platform system in which the other part is an online training log site. The site can be used to store and manage exercise logs and to customize the device.

To form an understanding of the extent of customization with this type of device we first identified all the features or items that could be customized on the device. Altogether 28 different features were identified as customizable. However, some features (for example the seven items related to personal settings, e.g. body weight) might not be considered as customization items as such, but rather as essential features to adjust according to the user's characteristics in order for the product to work as intended. Therefore we used the following criteria in the selection of items to be included in the study: the items should concentrate on the top-level customization that are central to the product use, the items should represent the main activities of customization based on previous research [8, 10] and at least some of the items should be customizable both on the device and on the online training log site. Due to the self-report method, it was important that it was reasonable to assume the users were aware of the customization options that were included in the questionnaire. Consequently, we concentrated on features that are explained in the user guide of the product in chapters 'Customizing Your Product' and 'Adjusting Your Settings'. Based on these criteria, ten different measures of customization in three different categories (functionality, general preferences and appearance) were chosen to be included. Top-level customization in this study refers literally to the highest level of customization, namely to the degree whether customization was committed or not on the chosen categories and their main sub-categories. For example, the extent of customization within a default exercise mode (e.g. how many displays and which data fields were customized) is out of the scope of this study.

Customizing the Functionality. Customizing the functionality of the device can only be done on the online training log site. The customization of the functionality of the device happens on two levels: first by choosing what exercise modes are stored on the device and second, by determining what is being measured and shown on the screen during the particular exercise. Consequently, this type of customization defines (part of) the content, and affects the interface of the device when using it during exercising. For example, less exercises stored in the device improves getting access to the functionality, and less screens attached to an exercises decrease the need for scrolling and therefore improves usability.

As a default there are eight different exercise modes stored in the device: alpine skiing, cycling, indoor training, mountaineering, running, trail running, trekking, and other sports (see Figure 2 for a screenshot of customizing the mode for running). The default modes can be used as they are, or they can be customized in the online training log, after which they need to be downloaded on to the device again. In addition, new modes for new exercises can be created from scratch and later modified if needed. In order to measure the top-level customization of functionality we asked if the user had:

- Edited displays to default exercise modes
- Added displays to default exercise modes
- Deleted default exercise modes
- Created new exercise modes
- Edited new exercise modes

Customizing General Preferences. Customizing general preferences was included because this type of customization can be performed both on the device and the online training log site. In addition, customizing general preferences has been found to be one of the main customization activities [10]. Of the options for general preferences, customizing formats of time, date and units was included in this study. Customizing the time and the date affects how the display looks, and customizing the unit system affects how the data is displayed while exercising. As a default, the device had as a time format 24 h, as a date format dd.mm.yy and the unit system was set to metric. Thus, if the user was satisfied with these formats, no customizing was needed. In the study, the user was asked if he had:

- Customized time format (12h or 24h)
- Customized date format (dd.mm.yy or mm/dd/yy)
- Customized unit system (metric, imperial or advanced)

Customizing the Appearance. According to previous research customizing the appearance is an important type of customization and is deeply rooted in the use of certain products [13]. However, with the device, there are limited possibilities for customization: the display can be inverted from dark to light, and vice versa (both on the device and on the online training log) and the strap can be changed. Customization of the appearance of the device was measure with the two measures:

- Invert the display
- Change of strap

3.2 Customization Questionnaire

The study was carried out as a part of larger research project investigating long term user experience during six months of product usage. During the research project, each month the respondents answered a questionnaire in English on their product usage and experiences with the product. The customization questionnaire reported in this paper was included in the third survey (third month into the study). Some basic demographic data (e.g. age and nationality) and data on sports activities collected at the recruitment phase of the study are included here. In addition, data on exercise habits, product usage and satisfaction with the product collected in each survey are combined with the customization data.

The 21 questions related to customization consisted of both closed and open-ended question, and were on the last page of the questionnaire. At top of the page, a description of customization was given. Multiple choice questions were used to find out about the extent of customization: the respondents were asked to mark from the list of ten customization items all the ones they had used during the total time of product usage. Closed-ended questions with a 7-point interval scaling (1-7; 7 signifying strong agreement) were used to evaluate the importance of customization, the ease of customization and the effects of customization (e.g. product being more useful and personal after customization). Open-ended questions were used to probe about the knowledge of customization of the product before taking it into use and the reasons for customizing the product. In addition, the respondents were encouraged to comment freely after questions.

3.3 Respondents

The respondents for the six month long study were selected among the customers who responded positively to the e-mail invitation to take part in the study by answering a short questionnaire with basic demographic, product purchase and usage questions. The invitation was sent to 521 registered owners of the product of which 190 (36%) expressed their interest to take part. 121 were chosen to take part based on three criteria: 1. Short usage time of the product; 2. Nationality; and 3. First come first served. As an incentive to take part, a respondent would receive a product of Suunto worth up to 269

USD after completing all six surveys.At the time of the data collection for this paper (third survey), the number of respondents had dropped to 110 (21 % of the whole sample getting an invitation). A further ten respondents and their responses were removed from the analysis due to missing data leaving the sample size to 100 respondents.

3.4 Analysis

For the quantitative data, nonparametric tests were selected due to the nonparametric (distribution-free) nature of the gathered data. The CHI Square was used in pairwise comparisons of categorical variables and Spearman's ρ was used in calculating correlations. A content analysis was performed to analyze the participants' descriptions of their main reasons for customizing the device. In the analysis, the basic needs of autonomy, competence and relatedness, and the motivations relating to them were used as the guiding framework [1].

4 Results

4.1 Characterizing the User

Demographic Information. The 100 respondents consisted of 93 males and 7 females, and their age varied between 22 and 64 years, with the mean age being 41 years (two modes: 37 and 45 years). 49 respondents had owned the device for three months and 51 respondents for four months. Over half of the respondents were from Europe (58 respondents), but the sample consisted nationalities from almost every continent: North America 18, Oceania 11, Asia 10, and (South) Africa 3. There were 17 different nationalities among the respondents. The four biggest nationality groups were Dutch with 13 respondents, Americans with 12 respondents, Finnish with 10 respondents and Australians with 9 respondents. 34 respondents have English as their mother tongue.

Sport Activities. Since the product is targeted for training and especially outdoors activities, we were interested in what activities the respondents engage in. From the list of 18 types of sports (with an open text for 'other' option), the respondent marked each sport he does. The respondents are active: on average they engage in seven different sports (range 2 – 16, mode 6), and altogether 46 different sports were mentioned. The most often mentioned sports were hiking (81), road running (70), trail running (60), road bicycling (60), going to the gym (49) and mountain biking (42). Thefavorite sport was asked with an open-ended question. Running was the favorite sport with 47 replies (includes 27 replies mentioning trail running specifically). Other top three sport activities were hiking and biking (off road, mountain or road) both with 14 mentions and skiing (off-piste, alpine or Nordic pole) with 7 mentions. On average, the participants exercise 4.3 times per week (STD 2.08, median 4, mode 3,).

Sports Watch Usage. The respondents were asked how often they use the device as a daily watch and in exercising (Table 1). Although the device is mainly targeted for use during different exercising activities, over half of the respondents (56) use it daily as a daily watch.

Table 1. Frequencies of the respondets using the watch in exercising and as a daily watch

		Using the watch in exercising				
		Less than once week or not at all	Once a week	Several times a week	Daily	Total
Using the watch as a daily watch	Not at all		2	3		**5**
	Less than once week or not at all	1		7	1	**9**
	Once a week		1	10		**11**
	Several times a week		1	13	5	**19**
	Daily	2	5	33	16	**56**
	Total	**3**	**9**	**66**	**22**	**100**

However, diverse use of the device is common: majority of the respondent (67 respondents) falls under a usage profile in which the watch is used at least several times a week both in exercising and as a daily watch. 16 participants use the device daily both as a time piece and in exercising. On the other hand, there are respondents who tend to use the watch more specifically for exercising: 21 respondents, who use the watch several times a week in exercising, use it only up to once a week as a daily watch. In addition, there are 8 respondents who mainly use the watch as a daily watch and only up to once a week in exercising. Furthermore, 4 respondents use the watch rarely in either use on a weekly basis.

4.2 The Extent of Customization

Table 2 gives the percentages of the 100 respondents that reported having customized that particular item. The most popular item to be customized was to invert the display. Almost as popular was the functionality related customization of editing displays to default exercise modes. However, overall, the rate of customizing was high with every item with the exception of customizing the appearance by changing the strap which was done by only one respondent.

Table 2. The percentage of the respondents who customized the particular item

Customization type	Item	Percentage
Appearance	Invert the display	83 %
Functionality	Edit displays to default exercise modes	82 %
Functionality	Add displays to default exercise modes	74 %
Functionality	Create new exercise modes	68 %
Functionality	Delete default exercise modes	58 %
General preferences	Customizing Time format (12h or 24h)	56 %
General preferences	Customizing Unit system (metric, imperial or advanced)	51 %
Functionality	Edit new exercise modes	48 %
General preferences	Customizing Date format (dd.mm.yy or mm/dd/yy)	43 %
Appearance	Change of strap	1 %

Customization Score. The customization score represents the number of the customization features the respondents reported having used. On average, the respondents had used 5.6 (STD 2.2, median 6, mode 7) of the ten customization features measured in this study. The high mean score indicates relatively comprehensive use of the features available, and may suggest general interest to customize in this context. However, the wide range of customization (range 0 – 9 items) and high variability across items measured among the respondents suggests there are differences in customization behavior. Grouping the respondent based on the number of items customized show there are four main groups (Table 3) that can be divided to low customization group and high customization group with the average customization score as a divider.

Table 3. Groups of respondents based on the number of items customized

Customization group	Number of items customized	Frequency
Low customization	0-3	19
	4-5	27
High customization	6-7	34
	8-9	20

Earlier studies have found differences in customization extent, for example, originating from gender [8] and the presence of dispositions [9]. The effect of gender was impossible to measure in this study due to small amount of females in the sample. However, user related dispositions 'Frequency of Use of Product' and 'Knowledge of Customization', and system related disposition 'Ease of Customization' were taken into account in this study. A Chi Square test was performed to determine if there was a difference between usage frequency of the device in exercising and use as a daily watch and customization score (low customization/high customization group). The test indicated a significant difference with usage in exercising ($\chi2 = 7.19$, df = 2, $p < .05$). However, the test failed to indicate a significant difference with usage as a daily watch ($\chi2 = 2.74$, df = 3, p= .433). The results indicate that use of the device in exercising may urge customization. Furthermore, a Chi square test was performed to determine if there was a difference between knowledge of customization and customization score. The respondents were categorized based on their knowledge of customization in three groups: 'Being able to customize was one of the purchase decisions', 'I knew about customization before purchase' and 'I did not know about customization before purchase'. However, the test failed to indicate a significant difference ($\chi2 = 2.72$, df = 4, p= 0.605). Correlation (Spearman's ρ) was calculated to study the relationships between the ease of customizing the functionality and the customization score. There was a weak correlation between customizing the functionality and customization score (Spearman's ρ r=.209, $p<.05$). This correlation indicates that there is a relation between the ease of customization of exercise modes and customization behavior, but it is impossible to say if ease of customization caused more customization. It may as well be the opposite: by customizing more the respondents may have enhanced skills in customizing and therefore it is perceived as easy.

In the following paragraphs the use of each of the different types of customization will be discussed separately.

Functionality-Based Customization. Customization of functionality was measured by three items relating default exercise modes (editing displays, adding displays and deleting modes), and two items relating to new modes (creating and editing).

The possibility to customize the default exercise modes was used extensively: 87 respondents had utilized at least one of the customization options provided. 50 respondents had used all the three options given, 21 had edited and added displays, 6 had only edited displays, 5 had edited displays and deleted modes, 2 had deleted modes, 2 had only added displays and one had only added displays and deleted modes. 68 respondents reported having used the customization option to create a new exercise mode, and 48 had also edited the new mode. 38 respondents had customized all three default modes and created new modes. It was more popular to customize the default exercise modes than to create their own. A Chi Square test was performed to determine if there was a difference between the usage frequency of the device in exercising and the use of functionality customization items. The test with create new mode indicated a significant difference ($\chi 2(2) = 6.16$, $p < .05$) suggesting that the respondent who use the device more in exercising create more new exercise modes.

General Preferences. Formats were also actively customized: altogether 71 respondents had customized at least one of the formats. To be precise, 35 respondents had customized all three formats; 27 respondents had customized just one of the formats (time by 13 respondents, unit by 11 and date by three), and nine respondents had customized combination of two formats. Altogether eight different combinations to customize (or not to customize) the three formats were found.

There are different conventional ways to present formats in different countries and we expected respondents from countries where the default formats are commonly used not to customize them. A cultural analysis was performed due to the unexpected high number of respondents who customized the formats. Among the 19 different nationalities there were 10 from countries (origins of 51 respondents) in where the same as the default date, time and unit formats set on the device are commonly used. In eight cultures (origins of 37 respondents) partly same as default formats are used: for example in Australia the dd.mm.yy style of date format and metric system is used, but for the time, the 12h format is used. In some countries, like Canada and Philippines (origins of 7 respondents), different notations, especially with regard to date and time, are used concurrently depending on the language of the citizen. In contrast, in US (12 respondents) none of the default formats are used suggesting a need for customization of formats for American respondents in our study.

Based on the clarification of the conventions in different countries, we investigated the customization of each format separately per respondent by comparing the respondent's nationality to the reported customization of the formats. The base line was the assumption that a respondent would use formats customary to his culture. This analysis aimed at investigating the possible effect of culture on the reported customization of formats by revealing frequencies of respondents who 1) customized according to expectations based on cultural notation and 2) customized contrary to cultural notation. Some respondents were left out from the analysis due to uncertainty of commonly used notations in their countries of origin: British and Canadian respondents (14) were left out from the analysis of time, Canadian, Filipino, and South African respondents (10) were left out from the analysis of date, and British were left out from the analysis of units.

Table 4. Customization behavior based on cultural notations

	Date (n=90)	Time (n=86)	Units (n=92):
According to expectations	62%,	65%,	55%,
Against expectations	38%	35%	46%

The results indicate individual preferences differ from cultural conventions with approximately third of the respondents. The difference from cultural conventions is further highlighted when investigating the customization of the formats with the biggest nationality groups further. For example, five out of twelve of the Dutch respondents had done some customization: one had customized the date and time format, one the date and unit, two had customized the unit and one the time. Further, five out of ten of the Finnish respondents had customized all the formats, one had changed the time and one the date. This leaves only three who had kept the default formats that are the common notation in Finland. In addition, half of the American respondents (6/12) had customized all formats, two the date and time, one time and date, one time and unit, one the date, and one had not customized any.

To understand the reason for customizing the general preferences contrary to cultural conventions, we looked into the responses to the open-ended question asking reasons for customizing the device (Analyzed more thoroughly in section 4.5). Since the question was general, only nine respondents specifically mentioned reasons to customizing the formats. Nevertheless, the responses show that convenience in a context or in use situation and willingness to experience and try out new things may trigger the customization of formats:

> *"I change the unit system for pacing a friend during a 100miler in the USA. I used the imperial system instead of the metric so we didn't have to do the conversion all the times" (ID64, Canadian)";*
> *"I liked trying different formats and using US formats vs European (…)" (ID43, American).*

Customization of Appearance. There were only two possibilities to customize the appearance of the device: by changing the strap and by inverting the display from light to dark or vice versa. The latter was the most often customized item.

Only one respondent had changed the strap. Analysis of the free text comments showed that many respondents were not aware of the possibility to change the strap but would be interested in it. On the other hand, there was couple of respondents who were not interested in customizing the appearance at all:

> *"No need to customize look. This one is perfect." (ID91).*

4.3 Reasons to Customize

In the analysis of qualitative data on reasons to customize, two responses were excluded due to answers relating more to the need for new customized features than reasons to customize. The analysis was based on Oulasvirta & Blom [1] who suggest the Self-Determination Theory (STD) [15] as a framework to understand customization. The reasons were categorized into the motivations relating to the attainment of three basic needs, autonomy, competence and relatedness.

Four respondents claimed they have not engaged in customization (yet) or that they are happy with their settings. These four respondents were on the low customization group (0-3 items customized). The responses from the remaining 94 respondents gave 157 reasons to customize: each reason given was categorized and thus, the average amount of reasons per respondents was 1.6. The results are presented below first with Figure 3 presenting frequencies of the reasons as motivation categories. After that the main findings on each motivation is described separately with quotes from the respondents. A category "Looks" was included in the analysis outside the analysis framework since responses relating to the improving the appearance did not fit to any of the categories.

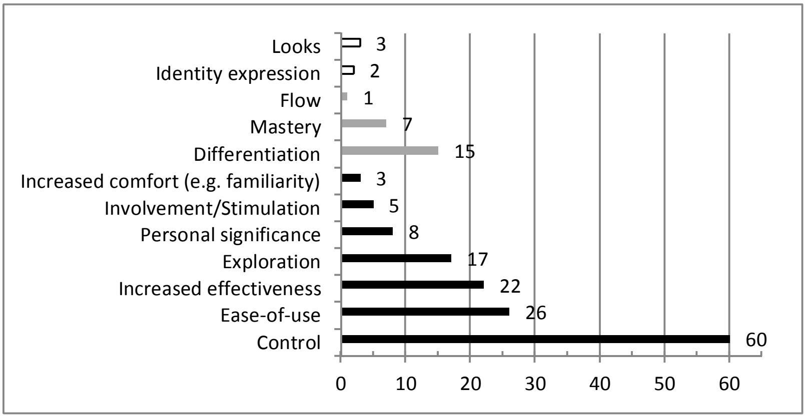


Fig. 2. Reasons to customize categorized in motivation types relating to autonomy (black), competence (grey) and relatedness (white). The graph presents number of reasons given by the respondents for each motivation type

Autonomy. The given reasons to customize were clearly most often related to control and the motivation relating to autonomy. This was highlighted by the use of wordings like "I want", "I need", "I like to have". The feeling of control was often related to ease-of-use and/or effectiveness of use. In addition, respondents often described the specific needs in relation to their activities: the importance to see the relevant information easily in a glance and without the hassle of going through different screens. At the same time, they recognized customization as "mandatory" in order to make the device personal. Some respondents engaged in customization just for fun or to try out.

"No company could make a watch which suits everybody, so customizing is mandatory for me. Nobody other than me knows what I want." (ID21)

"The main reason is that I want to decide what information is important to me while I exercise. I also want to control what buttons and how often I have to push a button in order to get the information I need." (ID54)

"Just to try out how it works, to have kayaking in the menu, and for having specific details on the watch display for different exercises to make it easier to use." (ID36)

Competence. Motivations to customize relating to the basic need of competence are to do with acquiring skills, differentiating oneself by preferences and interests and engaging in action. Often the competence motivations were mentioned along with control and the benefits of customization was acknowledged. In some responses the ability to customize or to learn *from* the act of customization was seen as beneficial to the individual.

"I deleted different sport profiles that I knew I would never, or rarely use. The same goes for the custom screen displays - I only display the parameters that I know will benefit me during training." (ID90)

"Also, I have a better understanding of the functions available and have been to fit the best functions to certain exercise types." (ID107)

Relatedness and the Looks. Customization of the device was clearly not driven by motivations of emotional or identity expression. There were many who claimed the reason for customization was to make the information provided by the device more personal but only two respondents claimed they customized the device (and not only the information) to make it their own. Three respondents specifically said they had changed the display color to improve the looks although most respondents had used the feature.

"I have changed to the dark display because it looks better. I have customized the exercise mode to also allow me to measure other activities and display I like to have while I am performing said exercise. I set the time format to military time (24 hrs) which is my preferred way to tell time." (ID77)

"Needed to customize the watch to make it feel more personal, to make it feel like it belongs to me. To make the intepretation of information from the watch easier." (ID106)

Summary. Both the low and high customization groups expressed for the most part similar reasons to customize. The main differences with them were that in the high customization group's responses control appeared more, whereas the low customization group mentioned motivations relating exploration more often.

4.4 Importance and Effects of Customization

Importance of customization and effects of customization (customization makes the product more personal and useful) was measured by asking an agreement to corresponding statements with a 7-point scale (1-7; 7 signifying strong agreement). In addition, the respondents had possibility leave comments on a free text field. Correlations

(Spearman's ρ) were calculated to study the relationships between the statements and customization score. Table 6 presents the basic descriptives and correlations with customization score. Generally, the respondents showed high agreement with all statements. Regarding the importance of customization, the results suggest it is more important for the respondents to customize the exercise modes than the basic functionality or appearance. In addition, the ability of customization to make the product more useful was appreciated:

"I like to customize my gadgets for my preference. It's like new every time I customize it." (ID101)

There was a positive correlation between all statements and customization score. This correlation indicates that respondents who expressed stronger agreement with the statements had customized the product more.

Table 5. Basic descriptives and correlations of customization score with statements relating to effects of customization

Statement	Descriptives	Cust. score
Being able to customize the exercise modes is important to me.	6.7 (STD.63, median 7, mode 7)	.340**
Being able to customize the formats of time, date and units is important to me.	5.9 (STD1.8, median 6, mode 6)	.328**
Being able to customize the looks of the product is important to me.	5.1 (STD1.7, median 5, mode 7)	.228*
Being able to customize makes the product more useful.	6.7 (STD.68, median 7, mode 7)	.337**
I feel the product is more personal after customization.	6.5 (STD.80, median 7, mode 7)	.347**
**. Correlation is significant at the 0.01 level (2-tailed), *. Correlation is significant at the 0.05 level (2-tailed).		

4.5 Satisfaction

Satisfaction with the device was measured with a simple question: "How satisfied are you with your Suunto Ambit?" with a 7 –point scale (1-7; 7 signifying strong satisfaction). The respondents reported very high satisfaction with an average score of 5.97 (STD 1.2, median 6, mode 6). Correlation (Spearman's ρ) was calculated to study the relationships between satisfaction and customization score but no correlation between the measures was found (Spearman's ρ r=-.013).

5 Discussion

The current results show the users of a sports watch customize their device extensively albeit a high customization group and a low customization group could be distinguished. There was only one respondent who had not customized his device at all and only one customization option (appearance-related: strap change) was used by only one respondent. However, little use seemed to be due to unawareness of the respondents about the option, and not because it was not appreciated. Customization of appearance (invert the display) was the most often customized item followed by the four main items of functionality customization. Functionality-based customization was rated as the most important, and customization was seen as an instrument that makes the product more useful. The respondents seemed to have practical reasons for customization: those who used the device more in exercising had also customized it more extensively. However, we cannot say if more customization resulted in more use (in exercising), a phenomenon seen with smartphones [8], rather, the product may be contrasted to a work tool: in Page et al [10] study on a word processor persons who used the software most, had customized most. This was also the case here: the respondents who used the device more in exercising, had customized it more. However, the finding does not allow proposing a cause-and-effect relationship between the variables. Rather, it may point to the fact that using the device more in exercising, the user becomes more aware of his needs with regard to measuring his performance and capabilities of operating the device during the particular exercise (e.g. contextual factors). Therefore, more needs for customization arouses, and more customization on the device is likely to be performed. The sport a user does may have an effect on customization needs.

In previous research the emphasis has often been on the customization of the appearance (e.g [9, 13]) because customization of appearance is the way to make the device look and feel "own" [6]. The results of this study suggest that customization of appearance is not the only condition to feel that a product is personal. Being able to customize the main functionality of the products, to better fit it to the individual needs (in an activity important for the individual), can bring about the feeling of the product being personal. In fact, the findings by Marathe & Sundar [11] on the effect of customization on sense of control seem to fit here: having control is a way to express one's identity. Furthermore, it the results indicate that the respondents got a product with looks they liked and thus, customization of appearance was not so important at least in order to express identity.

Control was the main motivation to customize with other motivations relating to autonomy following. No major differences between the respondents in different customization groups were found, but high customizers expressed some more control, whereas low customizers expressed more motivations relating to exploration. Overall, the results suggest that customization can be explained by the three basic needs, autonomy, competence and relatedness, with motivations in this context deriving mostly from the need for autonomy. Customization features are valued due to the sense of control the user has by having the tool to adapt the product according to own preferences, wants and needs. This sense of control was seen as a reason to customize in every customization type, not only in functionality customization.

The respondents who did not customize or customized minimally raise the question on their reasons for not customizing. In line with the results from Tossell et al. [8], it was learned that non-customizers have reasons for their behavior: there was not a need for it, the options were not suitable or the user was not aware of the possibilities. We feel the reasons for non-customizing may point to directions to further develop the customization features or, at least, improve the information about them.

Finally, we did not find correlation between satisfaction and customization extent despite it seems logical. A reason for this may be the overall high satisfaction with the product, and the insight that a user does not need to customize extensively to be satisfied with a product as long as his personal needs will be fulfilled and he can feel to be in control. Furthermore, a small matter, like the inversion of display color may be enough for a pleasurable experience.

Design Implications. Customization is a product feature that enables differentiation from competitors. However, available options should be easily found and recognized by the users or they may be unused despite their interest in using them. It seems there cannot be too many options: the more the better. The user should have the full control in deciding what is customized, but recommendations, for example, relating to contextual factors might be in order.

Limitations and Future Studies. There are certain limitations with this study. First, as the data was collected with an online questionnaire we cannot be sure how seriously the respondents took their reporting task, and since no observation was possible, the results are solely based on self-reports. In addition, it is not known if the respondents fully understood the concept of customization. On the other hand, the method seems appropriate for studying the top-level customization, and respondents with missing data were excluded from the analysis. Another limitation relates to the respondents that were drawn from the company's customer base. A customer interacting with a company might avoid giving negative feedback especially in a study with an incentive. However, the respondents were informed that the study is a research project conducted in a university and for example, all the communication with the respondents was from the university. In addition, reaching the respondents would have been quite difficult without the help from the company. Finally, the study reviewed the top-level customization of a sports watch: if more features were included to be measured, there might be more variety in the customization profiles.

In future studies, the aim will be to explore the true effect of customization on satisfaction and user experience by collecting data with repeated measures from the first usage (before any customization is performed) into couple of months use. With repeated measures the change in satisfaction and user experience could be compared to the extent of customization. At the same time, the process of customization could be explored: how often is customization performed and why, are the customized features actually used, and what kinds of effects arise from the process. Similarly, the preferred ways to do customization could be investigated. Another point of direction could be to study customization of a couple of products from the same user to explore the customization behavior and gain understanding in customization and motivation to customize in different context.

6 Conclusions

This study gives insight into customization in a new context. Previous studies have for the most part concentrated on the customization of appearance; this study introduces functional customization as an important part of product experience. This study explores the use of customization features of sports watch in order to assess the extent of customization, and identify reasons for customization in this context. Survey data from 100 users of a sports watch were analyzed to understand how they use the different customization features: general preferences, functionality and appearance. The findings show that in this context customization was extensive, but a high and a low customization groups were found. Despite the differences in customization extent, the users state similar reasons for customization: control, ease-of-use, increased effectiveness, and better fit to personal preferences. The motivation to customize in this context is for the most part related to autonomy: to the sense of control the user has by having the tool to adapt the product according to own preferences, wants and needs.

Acknowledgments. The author thanks Suunto and especially Tiina Taskinen for cooperation and enabling of the study, researcher Tiina Koponen (Human-Centered Technology, Tampere University of Technology) for her invaluable contributions to help carrying out this study and assistant professors Timo Partala and Satu Jumisko-Pyykkö (Human-Centered Technology, Tampere University of Technology), for their comments on the manuscript. This research was supported by the Finnish Doctoral Program in User-Centered Information Technology (UCIT) and the Finnish Funding Agency for Technology and Innovation (Tekes) through the project DELUX.

References

1. Oulasvirta, A., Blom, J.O.: Motivations in personalisation behaviour. Interacting with Computers 20, 1–16 (2007)
2. Weightman, D., McDonagh, D.: People are doing it for themselves. In: Proc. DPPI 2003, pp. 34–39. ACM Press, New York (2003)
3. Sunikka, A., Bragge, J.: Applying text-mining to personalization and customization research literature – Who, what and where? Expert Systems with Applications 39(11), 10049–10058 (2012)
4. Sung, J., Grinter, R.E., Christensen, H.I.: "Pimp My Roomba": designing for personalization. In: Proc. CHI 2009, pp. 193–196. ACM Press, New York (2009)
5. Ansari, A., Mela, C.F.: E-customization. J. of Marketing Research XL, 131–145 (May 2003)
6. Häkkila, J., Chatfield, C.: Personal customisation of mobile phones: a case study. In: Proc. NordiCHI 2006, pp. 409–412. ACM Press, New York (2006)
7. Ventä, L., Isomursu, M., Ahtinen, A., Ramiah, S.: My phone is part of my soul. In: Proc. UBICOMM 2008, pp. 311–317. IEEE Computer Society (2008)
8. Tossell, C.C., Kortum, P., Shepard, C., Rahmati, A., Zhong, L.: An empirical analysis of smartphone personalisation: measurement and user variability. Behaviour & Information Technology, iFirst, 1–16 (2012)
9. Blom, J.O., Monk, A.F.: Theory of personalization of appearance: why users personalize their pcs and mobile phones. Hum.-Comput. Interact. 18(3), 193–228 (2003)
10. Page, S.R., Johnsgard, T.J., Albert, Allen, C.D.: User customization of a word processor. In: Proc. CHI 1996, pp. 340–346. ACM Press, New York (1996)

11. Marathe, S., Sundar, S.S.: What drives customization?: control or identity? In: Proc. CHI 2011. ACM Press, New York (2011)
12. Blom, J.: Personalization—a taxonomy. In: Ext. Abstracts CHI 2000, pp. 313–314. ACM Press (2000)
13. Monk, A.F., Blom, J.O.: A theory of personalisation of appearance: quantitative evaluation of qualitatively derived data. Behavior and Information Technology 26(3), 237–246 (2007)
14. Suunto, `http://www.suunto.fi/us/en/products/Outdoor_Sports_Instruments/suunto-ambit/suunto-ambit-black-hr`
15. Deci, E.L., Ryan, R.M.: The "what" and "why" of goal pursuits: Human needs and the self-determination of behaviour. Psychological Inquiry 11, 227–268 (2000)

Online Requirements and Portal Design for Female University Science and Technology Students in Kenya*

Stephen Kimani, Eunice Njeri, and John Njue

Institute of Computer Science, Jomo Kenyatta University (JKUAT),
P.O. Box 62000 - 00200 Nairobi, Kenya
{stephenkimani,njeri.eunice,njuejohn}@gmail.com

Abstract. In science and technology (S&T) career progression, women drop out at virtually every step until extremely few are found in positions of influence and recognition. Although this is a global problem, it is even more critical in Africa. The number of female university S&T students in Africa has remained low for the past about three decades. There is thus a dire need to encourage/support such students especially in African countries, including Kenya. Online portals have the potential to motivate/support Kenyan female university S&T students. It is however critical to understand the characteristics and requirements of these students. While there exists literature on the differences between online female users and male users, women in S&T in African countries (including Kenya) face rather unique challenges. This paper reports a study that indicates the characteristics and online requirements of Kenyan female university S&T students, and then describes a corresponding online portal design.

Keywords: Online user requirements, user interface design, online portals, science and technology, female students.

1 Introduction

In science and technology (S&T) career progression, women drop out at virtually every step until extremely few are found in positions of influence and recognition. Although this is a global problem, it is even more critical in Africa [1]. Kabeer and Magnus [1] go on to say that in Africa unlike other regions of the world, the number of women pursuing S&T studies at university has remained low and stagnated for the past about three decades. According to Karanja in [2], the relatively lower number of S&T female students in Kenyan universities is attributed to factors such as: gender stereotyping; early pregnancies, heavier domestic workloads, etc. There is thus a dire need to encourage and support women who are pursuing S&T in universities especially in African countries, including Kenya.

* Research funded by Strengthening Information Society Research Capacity Alliance, SIRCA II.

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 403–410, 2013.

On the same note, research suggests that Internet-based interventions, such as online portals, can lead to positive changes in motivation, perception and behavior (e.g. [3], [4]). Online portals therefore have the potential to motivate and support Kenyan female university S&T students. One of the key prerequisites to designing appropriate online portals for Kenyan female university S&T students is to gain a deeper understanding of: their background, the challenges they face, their rating of potential interventions to the challenges, their Internet access and usage, their experience with and preferences for online portals. Various studies have been conducted in order to understand systematic differences between online female users and their male counterparts (e.g. [5], [6], [7]). While it is possible to glean from such studies the characteristics and requirements of online female users, it is however worth reiterating that women in S&T in African countries (including Kenya) face challenges that are rather unique from those faced by women in other parts of the world [1][2].

There exist a number of females and technology websites (e.g., wigsat.org, awis.org, etc), but they do not take into account the unique contextual challenges and specific needs of female university S&T students from Kenya. It should also be noted that gender-disaggregated data pertaining to women in S&T in Kenya is largely unavailable [8]. There is therefore the need to conduct a study that illuminates our understanding of the characteristics, needs, and online requirements of Kenyan women in S&T. This research work in particular characterizes Kenyan female university S&T students, describes their online portal requirements, and presents a corresponding online portal design. The contribution of this paper is therefore two-fold:

- It describes the specific online user requirements of Kenyan female university S&T students.
- It proposes and describes an online portal user interface design appropriate for Kenyan female university S&T students.

The rest of the paper is organized as follows: Section 2 describes the online requirements study in terms of the methodology, findings and emanating requirements; Section 3 describes the realized portal design; and Section 4 concludes the paper.

2 Online Usage and Requirements Study

This describes the study methodology, findings and emanating requirements.

2.1 Methodology

The study was aimed at determining online usage and requirements of Kenyan female university S&T students. We designed and administered a questionnaire targeting female university S&T students in Kenya. The questionnaire had questions on:

- Demographics of Kenyan female university S&T students
- Barriers and interventions related to Kenyan female university S&T students
- Internet access and usage of Kenyan female university S&T students
- Services and features that are desirable in online portals for Kenyan female university S&T students

The questions/items included in the questionnaire were informed by literature review on relevance of ICT in empowering women and promoting gender equality (such as [9][10]), common services/features in existing S&T portals, and also by preliminary interviews we had previously conducted with successful women in S&T in Kenya. Participants were recruited using: direct emailing, snowballing sampling, and class mailing lists. The questionnaire was made available online and in hardcopy. It was sent to 104 Kenyan female university S&T students. It was administered August–October 2012. Out of the 104 students contacted, 35 completed the questionnaire.

2.2 Findings and User Requirements

This subsection reports the findings and the emanating user requirements.

Demographics and Gender-Related Issues. Most of the respondents (86%) were undergraduate students. Female university students in S&T highlighted many challenges that they have personally faced as females students in S&T. The most highlighted challenge was *financial constraints*. The female university students also highlighted challenges that they know other female S&T students have faced. The most cited challenge was again *financial constraints*. From both the challenges that the students personally faced and the challenges they knew others faced, the top challenges were: *Financial constraints, Stereotypes related to women, Lack of mentoring, Academic programme is sometimes not relevant to the real-world, Lack of exposure, Lack of networks, Isolation/male-dominated environment, Discrimination, Lack of recognition, Family didn't have confidence in them, Lack of assertiveness*. The highlighted challenges echo what Kabeer and Magnus in [1] have highlighted as Africa-specific gender-related obstacles and what other researchers have highlighted as Kenya-specific gender-related obstacles. For instance: lack of access to good quality S&T education for females [1] (an obstacle related to *Lack of exposure*); failure of the formal S&T education to demonstrate the relevance of S&T to the females' day-to-day lives and challenges [1][11] (an obstacle related to *Academic programme is sometimes not relevant to the real-world*); and socio-cultural obstacles such as: gender stereotyping and customs, heavier domestic workloads for daughters than sons, etc [2] (obstacles related to *Stereotypes related to women and discrimination*).

As for the female S&T students' rating of interventions for addressing the challenges women in S&T face, the percentages can be seen in Fig. 1.

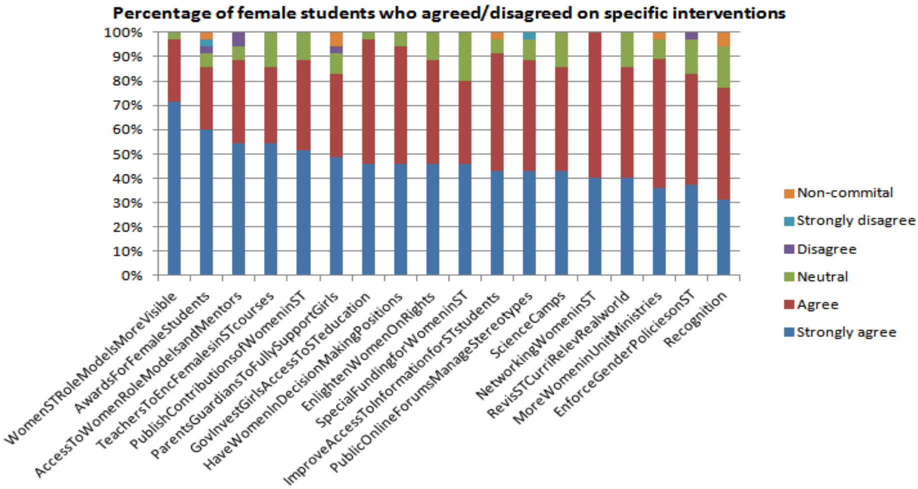


Fig. 1. Interventions for addressing the challenges

Internet Access and Usage. The main Internet access methods **in** and **outside** the university were: *wireless connectivity, Internet access through a personal airtime modem on laptop or computer, Directly from mobile phone, Internet café*. Most of the female S&T students reported a high frequency of their overall Internet usage (i.e. 71% use Internet several times a day). Most of the female S&T students use the Internet to *Send and receive emails*. More information can be seen in Fig. 2.

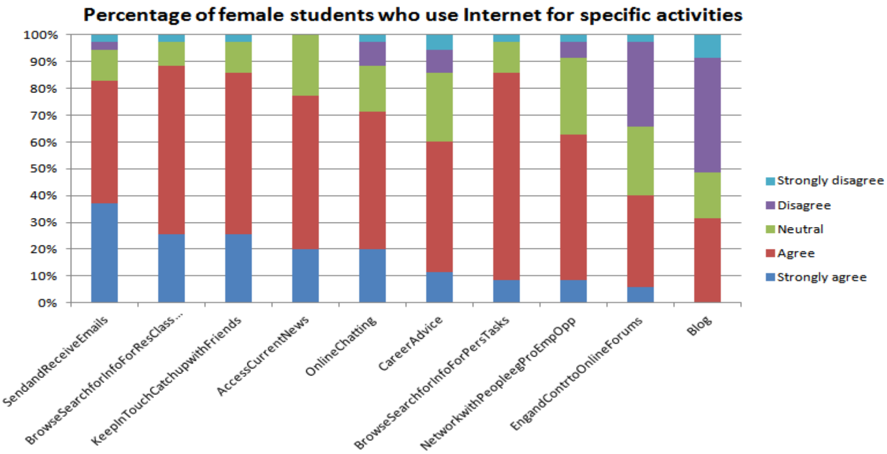


Fig. 2. Specific Internet activities

The female S&T students reported on the importance of specific Internet activities as seen in Fig. 3. *Browse and search for information for their research and class work* was leading and closely followed by *Send and receive emails*. It is worth mentioning that the Internet activities that female S&T students perform largely correspond to

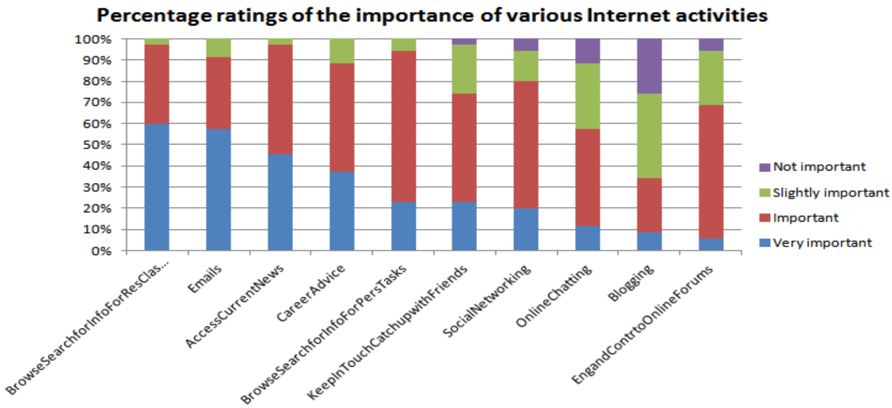


Fig. 3. Importance of specific Internet activities

those that they consider of high importance. For instance: *Email, Browse and search for information for research/class work, Access current news, Keep in touch/ catch up with friends, Career advice, Browse and search for information for personal tasks*.

Online Portal Services and Features. The female S&T students also reported on the importance of specific services and features if they were to be included in portals for female university students in S&T as seen in Fig. 4. We give a summarized description of the services/features that were rated as the most important in Table 1.

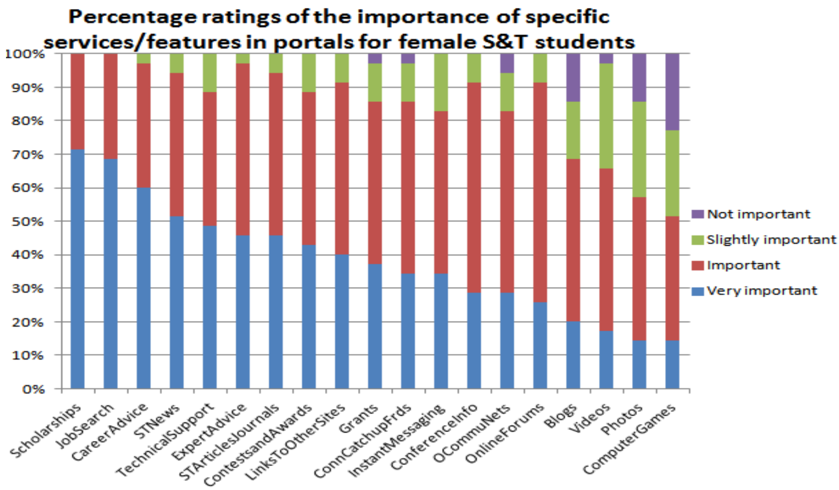


Fig. 4. Importance of services/features in future portals for female S&T students

Summary of the Findings and User Requirements. There were various specific portal services and features that were rated as important by respondents. The proposed portal should particularly seek to support the following services and features: *Scholarships, Job search, Email support, Career advice, S&T news, Technical support,*

Table 1. Most important services/features

Service/feature	Relationship with other findings
Scholarships	*Financial constraints* were earlier on highlighted as the challenge most of the female S&T students had faced, and knew other female S&T students who had faced it.
Job search	This is closely related with *Career advice*, which was also rated as an Internet activity of relatively high importance as a general Internet activity.
Career advice	As noted in the row above, this was also rated as an Internet activity of relatively high importance as a general Internet activity. It is also worth noting that this could be due to *lack of mentoring, lack of exposure* and *lack of networks*, which were previously highlighted as some of the top challenges that female S&T students face.
S&T news	This too had been rated as one of the most important general Internet activities.
Technical support	-
Expert advice	This could be due to *lack of exposure* and *lack of networks*, which were earlier highlighted as some of the top challenges that female S&T students face.
S&T articles and journals	It was earlier on reported that *Publishing the contributions of women in S&T* was one of the top interventions that the female S&T students had proposed.
Contests and awards	This could be due to *lack of recognition*, which was earlier on reported as one of the top challenges female S&T students face. It was earlier on also reported that *Awards for female students* was one of the top interventions these students had proposed.
Links to other sites	-
Grants	*Financial constraints* had been rated as the female S&T students' topmost challenge.
Connecting and catching up with friends	This was also rated as an Internet activity of relatively high importance as a general Internet activity. Moreover, *lack of network*s was earlier on reported as one of the top challenges female S&T students face.
Instant messaging	This is related to *lack of network*s, one of the female S&T students' top challenges.

Expert advice, S&T articles and journals, Contests and awards, Links to other sites (e.g. library catalogs), Grants, Connecting and catching up with friends, and *Instant messaging*. The proposed portal should also have services and features for addressing the main challenges that female S&T students face. Moreover, the proposed portal should have services and features for supporting the outstanding interventions for addressing challenges that women in S&T face. For instance: *information about financial/funding opportunities*; *online discussion forums* where female S&T students can manage discuss and educate one another on *stereotypes*; providing the *correct or accurate information about women*; *information about female role models or successful women in S&T*; enable female S&T students to *interact with mentors*; etc.

3 *StoreRooms* User Interface Design

In this section, we describe the preliminary user interface design of an online portal for Kenyan female university S&T students based on the findings of our study. The online portal is referred to as *StoreRooms* (*Science and Technology Opportunities and Resources Portal for Kenyan Female University Students*). The primary target users of the online portal are Kenyan female university S&T students. Toward ensuring that the proposed online portal addresses the requirements emanating from Section 2, the services/features are categorized/structured as seen in Fig. 5. Under each item in that level, we have the corresponding sub-items (e.g. Fig. 6 for sub-item *Opportunities*).

Fig. 5. Categorization and structure (high level layer)

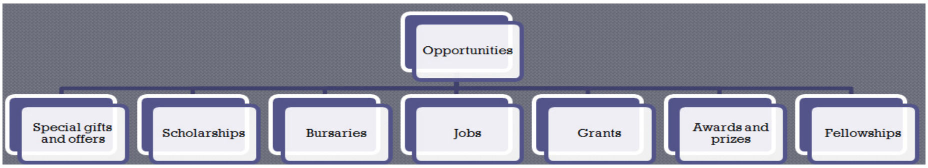


Fig. 6. Sub-items for menu item *Opportunities*

We have realized the preliminary user interface design based on the foregoing service and feature categorization/structuring. The corresponding design for the high level items is as seen in Fig. 7. An example of the design of the user interface for lower level layers can be seen in Fig. 8 (for *Scholarships* under sub-item *Opportunities*). Note in the lower level layers of the user interface, there is a home menu item to take the user back to the high level.

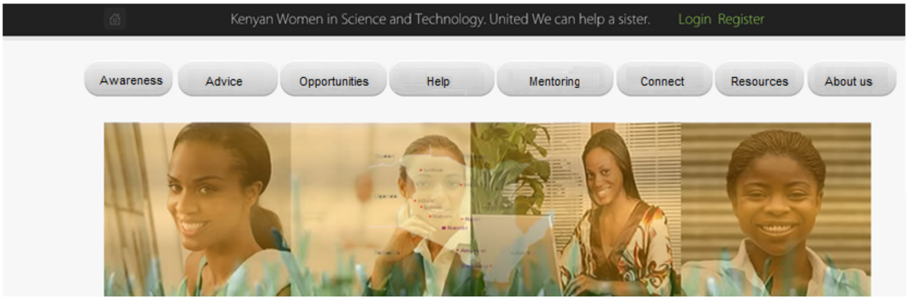


Fig. 7. Preliminary user interface design (high level layer)

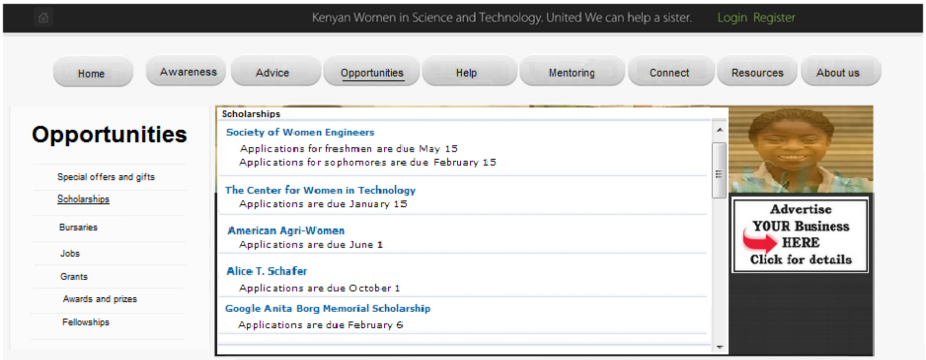


Fig. 8. Preliminary user interface design (*Opportunities* > *Scholarships*)

4 Conclusions and Future Work

In this paper we have described a study that was carried out in order to gain a deeper understanding of the characteristics and online requirements of Kenyan female university S&T students with regard to online portals. We have also described the emanating user requirements and the preliminary user interface design for an online portal intended for the Kenyan female university S&T students.

We noted in Sections 2 and 3 that the portal needs to have a service for supporting interaction between students and mentors. Based on that, we have consequently recently conducted a study involving potential mentors of Kenyan female university S&T students. We are analyzing the data collected from the potential mentors. The online portal is being developed using user-centered design. The current preliminary design will be subjected to heuristic evaluation. The results of the evaluation and the study with the potential mentors will be used to inform subsequent refinements of the design in order to realize an improved user interface design. After that we will conduct a user-based evaluation. Besides assessing the design, the user-based evaluation will investigate user access privileges and roles, and also determine strategies for user attraction and retention.

The main Internet access methods for the female S&T students suggest that these students rely largely on laptops and mobile devices for Internet access/use. It would thus be recommended to also consider developing a mobile version of the portal.

References

1. Kabeer, N., Magnus, E.: Building Women's Capacity in Science and Technology in the South - A Study for the Millennium Science Initiative/Science Institutes Group (2004)
2. Karanja, D.: Kenyan women hit out at male hold on science. SciDevNet website (2002)
3. Baghaei, N., Kimani, S., Freyne, J., Brindal, E., Berkovsky, S., Smith, G.: Engaging Families in Lifestyle Changes through Social Networking. International Journal of Human Computer Interaction 27(10) (2011)
4. Liebeskind, J.P., Oliver, A.L., Zucker, L., Brewer, M.: Social Networks, Learning, and Flexibility: Sourcing Scientific Knowledge in New Biotechnology Firms. Organization Science 7(4) (1996)
5. Odell, P.M., Korgen, K.O., Schumacher, P., Delucchi, M.: CyberPsychology & Behavior 3(5), 855–862 (2000)
6. Komathi, M., Ismail, M.: Influence of gender role on internet usage pattern at home among academicians. Journal of International Social Research 2(9), 308–318 (2009)
7. Joinson, A.N.: 'Looking at', 'Looking up' or 'Keeping up with' people? Motives and uses of Facebook. In: Proc. of CHI 2008, pp. 1027–1036. ACM Press (2008)
8. MoHEST: Gender Workplace Policy. Ministry of Higher Education Science and Technology. Republic of Kenya (2009)
9. Ramilo, C., Hafkin, N., Jorge, S.: Gender Equality and Empowerment of Women through ICT. Women2000 and Beyond. United Nations (2005)
10. Narayanaswamy, L., Sever, C.: Gender and ICTs. INBRIEF Bulletin 15 (2004)
11. Chetcuti, D.A., Kioko, B.: Girls' Attitudes Towards Science in Kenya. International Journal of Science Education 34(10), 1571–1589 (2012)

Finger on the Pulse: The Value of the Activity Stream in the Enterprise

Ido Guy, Tal Steier, Maya Barnea, Inbal Ronen, and Tal Daniel

IBM Research-Haifa, Israel
{ido,talst,mayab,inbal,taldan}@il.ibm.com

Abstract. The activity stream, which syndicates user activities across social media, has been gaining popularity on the web. With social media infiltrating the enterprise and higher portions of the workforce becoming accustomed to consuming information through activity streams, it also has the potential to play a key role in shaping the workplace. This work provides a first comprehensive study of an enterprise activity stream. We analyze different characteristics of the stream, its usage through a faceted search-based application, and the way users search it compared to traditional enterprise search. We also discuss various use cases of the stream, both from an individual employee's perspective and from an organizational perspective, exposing the potential value and role of the activity stream in the enterprise of the future.

Keywords: Activity streams, collaboration, cscw, enterprise, enterprise search, real-time search, social analytics, social business, social media, social search, social software, social streams, web 2.0.

1 Introduction

The *real-time web* [12] is characterized by highly intensive streams of updates and news generated by millions of users. Leading social media sites, such as Facebook, Twitter, LinkedIn, Myspace, and Google+, publish activity streams that include millions activities per day by users who write status updates, share links and photos, join groups, comment, and "like" others' activities.

Enterprise social media applications, enabling employees to share and interact behind the organization's firewall, have become popular in recent years and were extensively studied, from blogging [10] and microblogging [34], through wikis [4] and forums, to social bookmarking [25], social file sharing [31], collaborative task management [26], and people tagging [13]. Recently, several enterprise social media platforms have introduced enterprise activity streams, which syndicate employees' activities across the organization's social media. Examples include SalesForce Chatter(.com), SAP StreamWork(.com), Yammer(.com), and SocialCast(.com); however, the use and potential value of activity streams in the enterprise has yet to be extensively studied.

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 411–428, 2013.

The emergence of activity streams within the enterprise poses a great opportunity both from an individual (egocentric) and an organizational (sociocentric) perspective [14]. For the individual employee, this unique medium of highly intensive activities, concise in text and metadata, allows staying tuned with recent updates and discover new developments related to areas of interest. On the organizational level, the stream exposes timely information flowing at high pace from a wide variety of employees. As such, it can help increase awareness of projects and processes and expose recent trends and opinions to executives, administration people, or community owners.

In this work, we examine the potential value of the activity stream to the enterprise. We study our organization's activity stream, which syndicates activities across a wide variety of social media applications, including blogs, microblogs, wikis, files, bookmarks, and forums. All of these applications have been deployed in our organization for over three years and enjoy a large user base and frequent activity. Within our organization, an activity, defined as the basic unit of the stream, occurs at a frequency of less than four seconds during working hours. For instance, an activity in the stream can be: "*John Doe commented on the blog entry The Future of CRM in the Social Business blog*."

In our study, we used *Streamz*, an application that provides faceted search and analytics on top of an enterprise activity stream [18]. We analyzed the query logs of *Streamz* over a period of two months and interviewed 20 active users, in order to gain a better understanding of the activity stream usage scenarios from an egocentric perspective. Our analysis also focused on the comparison between search over the activity stream and traditional enterprise content search. We found that stream search is essentially different than "regular" enterprise search and serves for other complementary needs. Additionally, we augmented *Streamz* with a user interface that demonstrates the use of organizational analytics over the activity stream. Following the first set of interviews, we identified three lines of business (LOBs) that can benefit from such analysis: sales management, human resources, and IT. We conducted 8 additional interviews of users that belong to these LOBs to better understand their potential use of the stream from a sociocentric perspective. Overall, we believe our study provides a first comprehensive look at the potential use of the activity stream in the enterprise. Understanding this potential use can be fundamental to how employees collaborate and interact in the future workplace.

In the next section we discuss related work. We then describe the research setting, including a brief overview of the *Streamz* user interface and a description of the additional organizational analytics view. In the following section we describe a rich set of results, based on both quantitative and qualitative analysis, and then conclude.

2 Related Work

A large portion of the literature that relates to social streams studied the use of Twitter, the leading microblogging service (e.g., [22, 24]). Despite the vast popularity of research on Twitter, its search functionality has thus far received relatively little attention. Busch et al. [5] describe the architecture of Earlybird, the core retrieval engine behind Twitter's real-time search service. Naveed et al. [27] also examined search on Twitter, focusing on two challenges—content sparsity and the need for a static document quality measure. In 2011, the Text Retrieval Conference (TREC) initiated a microblog track that exposed a reusable test collection for the "study of Twitter as a real-time information repository" [28]. Teevan et al. [33] examined the use of search on Twitter and how it differs from usual web search. In our analysis, we compare search on *Streamz* with enterprise content search and link our results to their findings.

Microblogging services have also emerged in enterprises, used by employees to share messages about their work, promote new ideas, and converse around topics of interest. The use of microblogging in the enterprise has been studied in several papers. Zhang et al. [34] presented a case study of the use of Yammer, an enterprise microblogging tool, in a fortune 500 company, and showed that employees use it for a variety of work-related purposes, despite its "noise-to-value ratio". Ehrlich and Shami [11] compared microblogging inside and outside the workplace and found that internal microblogs were mostly used to solicit technical help or as part of a conversation, while the external microblogs were used for status updates and sharing of general information. Zhao et al. [35] explored microblogging's impact on informal communication at work and found that it may help colleagues know each other better and facilitate collaboration at work.

The Twitter stream principally consists of "tweets"—short status update messages of up to 140 characters. Our work focuses on a heterogeneous stream, in which status updates account for less than 10% of all activity. Other heterogeneous activity streams, such as the Facebook newsfeed, have been studied to a much lesser extent than Twitter. Paek et al. [29] used learning techniques to predict the importance of posts and friends on Facebook's newsfeed. They found that importance ratings varied greatly from one user to another and suggested personalization as a promising direction to address this issue. Other studies about the newsfeed of Facebook focused on privacy issues [3, 20] and diffusion models [32]. FriendFeed is another example of a heterogeneous activity stream that received some attention in the literature [7, 16], aggregating friends' activities across other social media sites.

A few recent studies examined activity streams in the enterprise, focusing on personalization techniques. Freyne et al. [15] proposed a method for narrowing the stream of the SocialBlue enterprise social network site based on person and action relevance inferred from users' browsing behavior. Daly et al. [8] suggested viewing the activity stream through "social lenses", based on user-defined collections of people and entities. Guy et al. [17] studied personalization of the stream based on a user model that includes people, terms, and places of interest. Our paper does not focus on evaluating methods for stream personalization, but rather provides a broad analysis to better understand the potential value of the activity stream to the enterprise.

We analyze the stream through the *Streamz* application, which provides faceted search on top of an enterprise activity stream and is described in our previous work [18]. Other studies have also proposed the use of faceted search to slice and dice aggregated feeds. Visual Backchannel [9] applied a faceted search approach on microblog conversations to show related people, images, and topics. Similarly to *Streamz*, they also use a stacked graph to present topics over time. The Eddi tool [2] represented topics as facets and suggested an alternate Twitter interface that allowed topic-based browsing. Hong et al. [21] introduced FeedWinnower, an enhanced feed aggregator that allows knowledge workers to filter feed items by topic, person, source, and time. Only an initial evaluation was provided, based on interviewing 15 employees, which pointed out two findings: all participants saw the value in faceted browsing and the most important facet was found to be the topic facet. In this work, we explore an activity stream of social media behind the firewall, rather than external feeds of websites or microblogging services.

3 Research Setting

3.1 Enterprise Activity Stream

We experiment with the activity stream of IBM Connections (IC) [23]—a social media application suite for the enterprise that includes eight types of applications, all of which have been deployed in our organization for over three years. IC publishes an activity stream of all public actions occurring across its applications. Table 1 details the different types of activities in the IC stream, including the originating applications, all possible actions, and frequency of occurrence as measured during one month of activity (percentage out of the entire stream).

3.2 The *Streamz* Applications

Our research is based on the *Streamz* application [18], which was designed to help users consume an activity stream. *Streamz* provides faceted search over the enterprise activity stream, as well as different types of analytics, including topic extraction, sentiment analysis, activity grouping, and personalization based on an interest profile.

Table 1. Activity types included in the IC stream

Application	Actions	%Occurrence
Blogs	create, edit, comment, like	4.09
Bookmarks	create	5.93
Files	create, edit, share, comment, like, download	14.53
Forums	create, reply	7.82
Microblogs	create, reply on own or another's board	9.45
Network	add to network, follow, tag self or another person	23.16
Tasks	create, edit, assign, comment, complete	8.25
Wikis	create, edit, comment, like	26.78

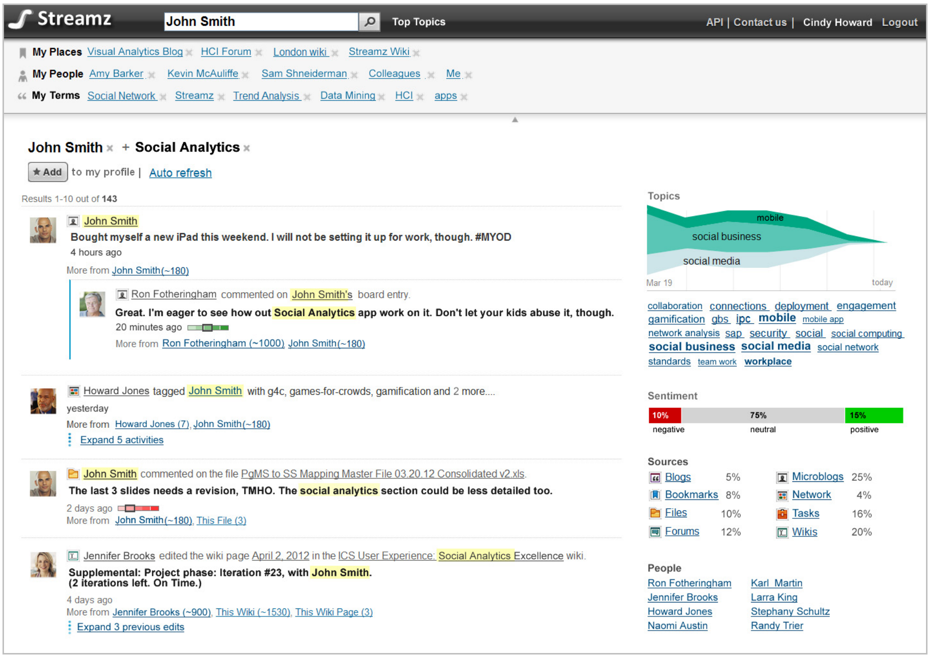


Fig. 1. Streamz main user interface

Main View. The main user interface of *Streamz* is shown in Fig. 1. The user can search the activity stream in various ways, but ultimately one user interface is used to present any subset of the stream. The UI consists of three main parts. Its main component, the stream (on the left), displays the activities in the stream based on the search criteria currently applied. Activities are presented in reverse-chronological order, with indication of their freshness (e.g., "2 days ago"). Each underlined entity within the activity description is a link to its corresponding IC page. The "more from" line allows the user to re-filter the stream to any of the entities in the current activity. The second component appears on the right and includes the facets [1], which summarize various aspects of the stream, including key topics, sentiment, source application, and active people. By clicking a facet, the user can drill down (refine) the results to the clicked value. The third component is the user's interest profile, which appears on the upper section for authenticated users. The profile includes people, terms, and places (wikis, blogs, forums, etc.) of interest and serves for personalizing the stream to a mix of activities according to the interests of the individual user [17]. The user can add profile items by saving search queries that yield interesting results.

Overall, a user of *Streamz* has four different ways to search the activity stream: (1) through a manual search query in the search box at the top of the page; (2) by clicking a profile item; (3) by clicking one of the "more from" links; and (4) by clicking a facet. The first three initiate a new query, while the latter refines the current results.

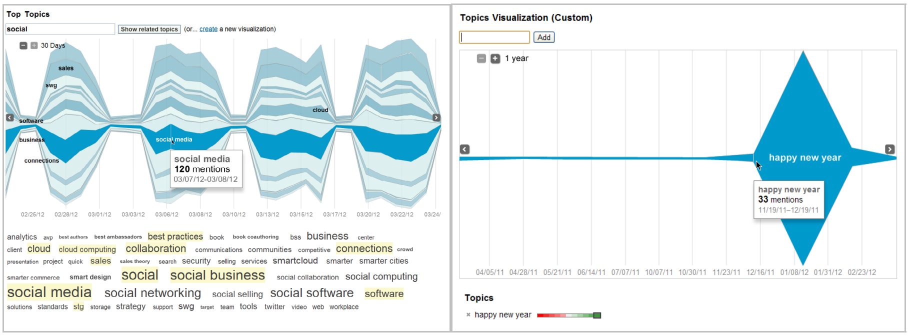


Fig. 2. Organizational views (Left:) Top topics view (Right:) Custom visualization view

Streamz uses various analytics components: (1) topic extraction, which associates the key topics with each activity; (2) sentiment detection, which associates each activity with a one of three sentiment classes: negative, neutral, and positive; (3) activity grouping, which collapses together similar or related activities. *Streamz* implements three types of grouping: a microblog *thread*, which includes a message and all its replies (see the first activity in Fig. 1); a *duplicate* activity, which includes consecutive wiki or blog edits that are identical but have different timestamps (see the last activity in Fig. 1); and a *compound* person tagging in case a user has tagged the same person with multiple tags (see the second activity in Fig. 1).

A detailed description of *Streamz*, including its user interface and analytics, is provided in [18].

Organizational View. For this study, *Streamz* was extended with a user interface that enables to further gain organizational insights from the stream as a whole. To this end, we added the top topics page, which can be reached by clicking the corresponding link at the top of the main view. The page presents in detail the leading topics in the organization, as extracted from the entire activity stream. The user can also enter a query to view the top topics that relate to that specific query. Fig. 2 (left) illustrates the top topics view. The upper part is a stacked graph visualization [6] of the top 20 topics over time. The user can change the time range from 30 days to 3 months or to a year. The time range is partitioned into 12 sub-periods. Hovering over a layer in the graph displays a pop-up window showing the topic represented by this layer, the sub-period corresponding to the current cursor position, and the number of activities that mention the topic during that sub-period. The user can move the cursor vertically to move between layers or horizontally to move across sub-periods. The word cloud below the stacked graph shows the top 100 topics that relate to the search query (or to the entire stream, if no query was entered). Topics that are part of the user's interest profile are highlighted. Upon clicking a topic, the user moves back to the main view, presenting activities that match the topic.

Users are also allowed to create their own custom stacked graphs by clicking the 'crate a new visualization' link. In this view, illustrated in Fig. 2 (right), the user can add and remove topics of her choice to an initially-empty stacked graph. A sentiment bar for each of the input topics appears below the stacked graph. This view is useful to track the evolvement of a single topic or compare a few topics over time.

3.3 Experimental Setup

Our evaluation is primarily based on an analysis of the *Streamz* query logs, which document every query issued to the system along with a timestamp, the user's IP address, and ID if they are authenticated. We analyzed the logs recorded during two months of activity. We also conducted interviews with 20 users who used *Streamz* during the selected time period to get an in-depth understanding of why and how they use *Streamz*. Interviews were semi-structured and covered general use of enterprise social media, usage scenarios of *Streamz*, and different aspects of search in *Streamz* (repeated queries, saved queries, and differences from traditional enterprise search). They were conducted by phone and lasted half an hour each. The interviewees originated from 11 countries and spanned different usage levels of *Streamz*. Three of the 20 interviewees were managers. Following these interviews, we identified three lines of business (LOBs) that can potentially benefit from the stream's organizational viewpoint, and followed up with 8 *additional Streamz* users who represented these LOBs. These interviews lasted 45 minutes, conducted by phone, and focused on the potential use of the stream to the LOB. Interviewees were presented with the organizational view user interface as a demonstration for how insight could be drawn from the stream as a whole.

4 Result Analysis

Our results are organized as follows. We open by a general descriptive analysis of the organization's activity stream along the inspected time period, to provide a sense of what the stream consists of. Since this is a new domain of research, we thought it is important to describe these characteristics as observed in our organization. We then describe the egocentric use cases of the enterprise activity stream as reflected in our initial interviews and through the usage analysis of *Streamz*' main UI. The following subsection inspects in detail the key functionality of searching the activity stream through analysis of the queries issued to *Streamz* along the time period. These are compared with the query log of a traditional enterprise search engine. The final subsection describes the organizational use cases as reflected through our follow-up interviews with LOB representatives.

4.1 Stream Statistics

Overall, the activity stream included 628,947 activities along the 60-day period. As expected in an enterprise environment, a sharp difference existed between the number of activities on a working day—13,285 on average (stdev: 3,937, median: 13,638,

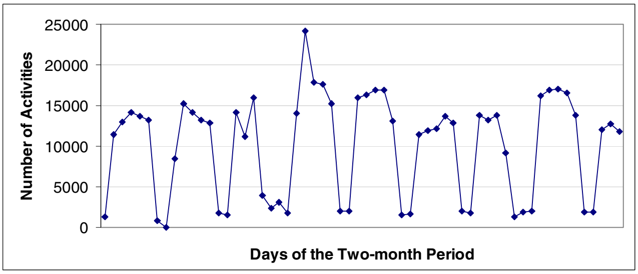


Fig. 3. Activity in the stream along the inspected time period

max: 24,181), compared to 1,735 on a weekend day (stdev: 630, median: 1,833, max: 3,089). Fig. 3 shows the volume of stream activity over the inspected time period.

As mentioned before, the rightmost column of Table 1 displays the occurrence frequency of activities per source application (percentage of activities out of the entire stream). Wikis were the most common source of activity, producing over a quarter of the activities in the stream. The fact that each wiki edit saved generates another activity inflates the number of wiki activities. Network activities and files were the next most common activities. Microblogs constituted only 9.45% of the stream, and blogs were the least common with only 4.09%.

The activities of the stream originated from 56,866 distinct authors and 51,144 distinct places. The top 10 authors, all with over 750 activities along the time period, originated from 9 different countries and 4 organizational divisions. The average number of activities per author was 11.07 (stdev: 57.34, median: 3, max: 1,340). Fig. 4 (left plot) shows the distribution of activities by author, which follows a power law with slope *α=-1.64*. The average number of activities per place was 9.65 (stdev: 36.43, median: 2, max: 1,964). The top 10 active places, all with over 1,000 activities, included 4 wikis, 2 blogs, 2 forums, and 2 tasks. Fig. 4 (right plot) shows the distribution of activities by place, which follows a power law with slope *α=-1.58*.

The average number of words per activity as presented on *Streamz* was 28.46, and varied substantially among activity types, from 8.53 for network activities to over 45 for blogs, forums, and wikis. File and bookmark activities were also concise, while microblogs and tasks were close to the average. After applying sentiment

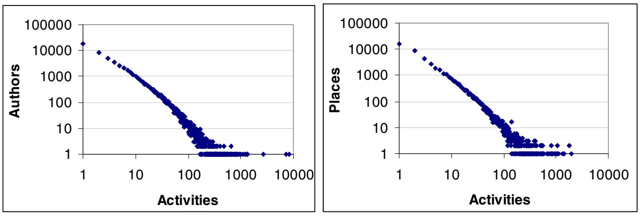


Fig. 4. Distribution of activities (Left:) over authors and (Right:) over places

classification, 16.43% of the activities were identified as positive, whereas only 0.74% as negative. One of our interviewees reflected: "*We have a very 'positivish' organization, it's a cultural thing […] we are implicitly encouraged to say good things and avoid saying bad things*." Inspecting negative activities, we observed that they commonly referred to technical issues or bugs. A smaller portion expressed a negative opinion about a topic, a technology, a tool, or a process.

The highest portion of non-neutral activities was detected for the wordy sources—forums, blogs, and wikis (about a third were non-neutral)—and for microblogs (28%). The other sources had 10% or less non-neutral activities, with almost 100% of the network activities being neutral. Overall, we found the set of activities in the stream to be diverse in length and sentiment. This may affect the way it can be used by different tools or systems. For example, an application that is mostly interested in sentiment analysis may only focus on activities that stem from microblogs, forums, blogs, and wikis, disregarding activities referring to bookmarks or to the network.

After applying our three grouping methods, the number of activities in the stream shrank to 83.15% of the original. Table 2 summarizes the statistics, including the number of grouped activities per each type, their average length, and their impact on reducing the number of activities. Grouping of duplicate activities was the most common type, leading to a reduction of almost 40% in the overall number of wiki activities. A smaller reduction was associated with blogs. Threading substantially shrank the number of microblogs, while people-tagging activities shrank by almost 60% due to "compound" grouping. The latter produced the highest length of grouped activities—almost 5 activities per group on average. Overall, we observe that by several basic rules, grouping can substantially condense the stream. All three grouping methods we experimented with produced an average group length of 3.5 or higher. Further research is required to validate that grouping renders a more readable stream, and to extend grouping to additional sources and methods, such as grouping by topic or by place (e.g., 3 people edited a wiki). Grouping can also play a key role when filtering the stream, since a grouped activity may attract extra interest.

4.2 Egocentric Usage Scenarios

In this section, we analyze the personal (egocentric) use of the activity stream as reflected through our set of 20 interviews and the use of the *Streamz* main UI along the inspected two-month period of our study.

Table 2. Characteristics of grouped activities

Type	Number	Avg Length	Source	Reduction
Duplicate	26,978	3.5	wikis	39.29%
			blogs	12.79%
Thread	8151	3.55	microblogs	35.39%
Compound	5263	4.91	person tags	59.93%

During the study period, 239 authenticated employees used *Streamz*. They originated from 26 countries and spanned the different divisions in our organization: 34.7% were from Software, 22.6% Services, 21.8% Sales, 9.2% Research, 7.1% Headquarters, and 4.6% others. Of these users, 14.2% were managers, which is similar to the percentage of managers in the entire organization. None of the top 10 users were managers. These results differed from those found for people search in the enterprise [19], where the majority of top users were managers. In contrast, it appears that activity stream search is as equally popular for employees as it is for managers. We note that this finding reflects the early adoption phase and may change with time.

Only 38.08% of the users used *Streamz* for more than one day, implying that the use of the activity stream in the enterprise is still in its early adoption phase. The average number of days per user was 2.46 (stdev: 3.08 max: 23). 28.57% of the *Streamz* users did not author any activity in the IC stream over the inspected time period. On the other hand, 18.07% were avid users of IC, who authored at least 100 activities during that period. The average number of authored activities by *Streamz* users was 79.43 (stdev: 172.77, median: 19.5, max: 1,340). Overall, users spanned a wide range of IC activity level, from non-active to highly active.

Many interviewees mentioned the term *dashboard* with regards to how they use *Streamz*. "*[Streamz is] useful as a dashboard view that provides a snapshot of the ongoing activities across a wide variety of applications in one place*," said one interviewee and another noted: "*It's an awareness system, monitoring what's going on, what are the things I need to know, what matters to me*." Another said: "*Similarly to Facebook, Streamz gives you a diverse feed of things you care about […] It really allows to keep your finger on the pulse of the organization*."

Another common usage scenario was getting a sense about what people say and think. "*I use Streamz to stay updated with what is being said about our projects*" said one interviewee. Another noted, "*Streamz gives me the 'talk' or 'vibe' about something going around […] understand who talks about it and where*." This benefit was particularly highlighted by salespeople: "*I use Streamz to understand what people think about new material we create for a sales opportunity. For example, we created a presentation deck and want to understand how much traction it gets*" noted one interviewee. Another salesperson said: "*I usually look for competitive information […] what people write about our own and others' products, what feature they want to see*."

Some interviewees mentioned that their interest in the stream revolves around individuals or groups, for example, "*I use Streamz to get a glimpse of what a group of people I'm interested in is doing,*" or "*I mainly look for my peers, my department, and my report chain to stay up to date*." Another interviewee revealed, "*When a new CEO was appointed and said she would use social media to communicate, I searched to see how many people started to follow her*."

Using *Streamz* to track specific topics of interest was also frequently mentioned. "*I use Streamz to see what's new in the technologies and subjects my work involves*" said one interviewee. Another told us: "*I work in the finance industry and I use Streamz every day to see what's new in the finance sector […] usually using three keywords: finance, economics, and banking*."

4.3 Searching the Activity Stream

The following section focuses on analysis of the search operation on top of the activity stream, in an attempt to understand if and how it is different from traditional enterprise search. 145 authenticated users (60.67%) issued at least one query through one of the four mechanisms that enable doing so in *Streamz*. On average, each of them issued 8.55 queries (stdev: 14.11, median: 3, max: 97), 5.75 of which were manually-typed search queries, 1.41 were issued by clicking on a profile item, 1.08 by clicking on a facet, and 0.31 by clicking "more from".

On average, 7.35 of the 8.55 queries were unique (85.96%). Teevan et al. [33] reported a substantially higher portion of repeated queries for both Twitter (55.76%) and the web (34.71%). The low percentage of repeated queries may reflect the fact that *Streamz* is not yet used so frequently as Twitter and web search engines. However, it is also likely that *Streamz* users take advantage of the ability to save queries to their interest profile, thus sparing the need to repeatedly search for the same things. This point was made clear by several interviewees. For example, one said: "*When my query produces interesting results I save it to my profile so I can keep track.*" Inspecting the profiles of *Streamz* users, 76.1% were non-empty, with an average of 11.2 saved queries per profile (stdev: 9.74, median: 9, max: 58), indicating that query saving was indeed in common use.

We next compare the queries issued to *Streamz* with queries issued to a more traditional enterprise content search engine, called *SaND* [30]. *SaND* is used to search for documents in the intranet, such as bookmarked pages, blogs, or wikis. It is not the official intranet search engine in our organization, but is used by several hundreds of employees every month. Alongside the document results, *SaND* presents related people, related tags, and source applications as facets, in a similar manner to *Streamz*. The similarity of both the usage levels and the user interfaces of *Streamz* and *SaND* enabled us to more neatly examine the differences stemming from the type of searched objects (activities versus documents) and the corresponding sorting of results (by recency versus by relevance).

To get a sense of the basic difference in result freshness between the two search applications, we inspected the overlap in the top 10 results for the top 50 queries common to both *Streamz* and *SaND* over one week. We found that the average overlap for *Streamz* across the 50 queries was only 14.2% (stdev: 8.6%, median: 10%, max: 60%), while for *SaND* it was as high as 92.4% (stdev: 11.6%, median: 100%). These differences demonstrate that within a week, most *Streamz* search results are replaced with new ones, while the vast majority of *SaND* results remain steady. This highlights the dynamic real-time nature of stream search, where results continuously change, and further motivates the comparison with traditional search.

Table 3 compares the characteristics of queries in *Streamz* versus *SaND*, based on the 50 most popular queries in each, as well as the top 50 queries common to both. Query popularity was determined by the number of unique individuals who issued the query at least once over the inspected time period. Teevan et al. [33] conducted a similar comparison between Twitter and web queries and found that Twitter queries were noticeably shorter than web queries (12 chars 1.64 words versus 18.8 chars 3.08

words, respectively). By contrast, in our results, *Streamz* queries were found to be somewhat longer than *SaND* queries. In general, queries in the enterprise appear to be shorter than queries on the web.

While a considerable portion of *Streamz* queries referred to individual people (18%), none of the *SaND* queries did. As already mentioned, one of the key usage scenarios for stream search is keeping track with the activities of a person of interest. Teevan et al. [33] also found a substantially higher portion of people searches in Twitter than on the web: 15.22% of the Twitter queries referred to celebrity names, compared to only 3.11% of the web queries. In our case, the people sought for are not celebrities, but rather coworkers or interesting employees in the organization.

On *SaND*, 36% of the queries referred to internal tools, such as a travel reservation system, a procurement system, a CRM system, and utilities such as an internal browser plugin, a survey tool, and a URL shortener. These queries are navigational in nature, i.e., they aim to get to a particular resource. The *Streamz* queries, in contrast, barely included references to internal tools. Teevan et al. [33] found a similar trend: while many of the web queries were navigational, Twitter queries were not.

A considerable portion of both the *Streamz* and *SaND* queries referred to internal products or projects (24% and 34%, respectively). Based on the comments we received in our interviews, it seems that the use of each application to search for projects or products is different. *SaND* is often used to find the main page of the product, or to locate its internal deployment. *Streamz*, on the other hand, is used to learn about latest features, opinions, plans, or issues that relate to the project or product. Another common query type to both *Streamz* and *SaND* referred to global topics or themes, such as “mobile”, “user experience”, “social media”, “big data”, and “travel”. Again, the use of each application to search for such topics seems to be different, as demonstrated by one of our interviewees: “*For a query like ‘cloud computing’, I would search SaND to get to the main wiki or forum, while I would query Streamz to get a sense of the recent buzz around it and understand how it can be relevant to my job*.” The list of popular common queries includes two main types: internal projects/products and topics/themes. As mentioned, the usage scenarios for these may often differ between content and stream search.

Table 3. Characteristics of *Streamz* queries versus *SaND* queries and queries common to both

	Streamz	***SaND***	**Common**
Avg length (chars.)	**10.38**	8.72	8.4
Avg length (words)	**1.4**	1.26	1.24
Is person	**18%**	0	2%
Is internal tool	2%	**36%**	2%
Is internal proj./prod.	24%	34%	**50%**
Is topic/theme	**34%**	28%	**34%**
Starts with #	**16%**	0	2%

The hash symbol was used in quite a few of the popular *Streamz* queries (16%), although a smaller percentage than found for Twitter queries (21.28%) [33]. This is likely because microblogs occupy only a subset of the stream, and the rest of the activities do not typically include hash tags. The three most sought-for hash tags were: *#socbiz* (referring to a theme), *#ls12* (business event)*, and #helpmeit* (a mark for the IT department). None of the popular *SaND* queries included hash tags.

Our interviewees shed more light on the differences between content search and stream search. Content search was said to be used for navigation to pages in the intranet of internal tools and projects and for finding the main projects or products related to a topic, or other 'stable' information, such as tutorial presentations, sales material, or marketing methods. On the other hand, *Streamz* was said to be searched for getting a sense of what people are saying right now about a topic, find something that was said before, get updated with recent news regarding a topic, a product, or a person, and understand the latest insights and buzz. One interviewee said: "*Streamz helps since it connects the content with the person who created it and the place it came from […] then, for example, I can locate people with similar problems.*" Another interviewee explained the difference when it comes to repeated queries: "*In SaND I repeat a query to re-find the same link, while in Streamz I repeat it to get the latest updates*." Another referred to the analogy of **searching** the content versus **filtering** the stream: "*[In Streamz] I don't search, I filter. Rather than trying to find something I need, I am narrowing the stream with different signals I remember such as people, keywords, and time range, in the hope to discover interesting updates*".

The value of the results' freshness in *Streamz* compared to content search was demonstrated by a few interviewees. One commented: "*I was looking for information on an annual organizational process and in SaND I found old pages since this process has been going on for many years. In Streamz I could instantly find the reference to the process taking place this year*". Another interviewee said: "*Streamz gives me the freshest information someone has bookmarked or wrote status about, while in SaND I often get old links sometimes even no longer valid*". Another stated: "*On SaND, the results are expected, while on Streamz I look to discover new things […]; often I find interesting stuff I did not even know I was looking for: a new interesting document, a new reference to my work, or even someone I know who got promoted and people congratulate him on his board*."

4.4 Organizational Usage Scenarios

31.8% of *Streamz* users accessed the organizational view. 83.1% of the interactions with the custom visualization view referred to addition of new terms to the visualization, while 16.9% referred to term removal. The most popular terms inspected through that visualization included the organization's name, its leading brands and divisions, prominent product names, broad area names ("social media", "cloud computing"), bursty topics (e.g., a company acquired by the organization), events (both one-time and recurring), people names, and source names ('wikis', 'forums', 'bookmarks').

Some of our interviewees saw the value of the top topics view from an egocentric perspective, for example: "*It gives me a company-wide view of what's going on, like* *reading the headlines of a newspaper*" and another described: "*I look at this page to see the main topics, what is hot and talked about […] I then select topics of importance that day to get a further look at the activities around those*". Another interviewee said: "*I use this view to search a person to see what he writes about the most and how his topics change over time*".

Through our interviews, we identified three specific lines of business in the organization that can benefit from the organizational view provided by the stream. We elaborate on these scenarios in the rest of this section and demonstrate each through quotes from follow-up interviews, conducted with four sales managers, two HR people, and two IT people, who used *Streamz* throughout the inspected period.

Sales Manager Scenario. Social media opens more channels for salespeople to document their activities, for example through blogs, microblogs, wikis, or shared files. One of our interviewees noted: "*Reporting on sales activity, such as customer meetings, through social media, especially microblogs, can serve as a good replacement for current CRM tools, since [the former] are much simpler and faster to use. This is especially important since sellers are usually on the road and need a simple interface they can use on their mobile device.*" Sales managers and executives now have the opportunity to learn more about the activities in their groups or divisions by analyzing the stream. Our follow-up interviews with four sales managers revealed several specific scenarios and requirements:

- "*I would like to be able to search for a customer name and see my group's volume of activity along a time-line, with a possible breakdown to each person in my group*"
- "*[…] use the 'top topics' view to see the activity in my department fragmented by key customers*"
- "*[...] know in real-time whether there is a sharp drop in activity or negative sentiment regarding a certain client*"
- "*[…] identify overlapping activities with the same customer […] We are a big company and often times we act in parallel, duplicating efforts and generating complexity without being aware*"
- "*[…] get actionable recommendations based on anomaly detection […] For example, create a sales opportunity around a specific trending topic that is relevant for a customer; or arrange a meeting with a client with whom activity has decreased*"

Human Resource Scenario. As more employees share their thoughts and opinions through social media, Human-Resource (HR) professionals can investigate the "employee voice" reflected in the activity stream to better understand the sentiment around certain policies, processes, or programs, and the main topics employees are concerned about. Activity stream analysis can replace traditional internal surveys and interviews, provide more authentic feedback, and allow taking action in near real-time to address emerging issues. We interviewed two HR practitioners who shed more light on how a potential HR dashboard can be built on top of the activity stream:

- "*A key desired capability is segmentation of the stream based on person attributes originating from HR systems, such as country, business unit, job role, seniority, band, salary, or performance grade*"
- *"[…] slice and dice the stream to view the sentiment and top topics in different segments of the organization, for example the sentiment of people with less than five years seniority in the Services division*"
- "*More important than getting the current status across the organization is getting the change map: where are the most drastic changes taking place right now?"*
- "*Mix the internal stream data with what employees say on external social media, which may reflect an even more authentic opinion […] Allow comparing between the two*"
- "*Analyze the flow of information and collaboration between departments and countries […] who is talking to whom and where do gaps exist?*"
- "*Alerts […] As opposed to just searching, I would like the system to push me notifications when a change happens, such as a new trend in South America or a drop in sentiment around BCG process for managers"*

IT Scenario. Discussions about organization's internal tools are quite popular on enterprise social media. Internal IT teams can use the activity stream to track new issues and enhancement requests, and get a view of the strengths and weaknesses of the tools they deploy. As mentioned above, in our general stream analysis, we discovered that many of the activities with negative sentiment referred to issues, bugs, or error notifications from internal tools. Our interviews with two members of the IT department revealed the following needs:

- "*[…] get a sense of the 'popularity' of tools by the number of activities that refer to them […] view popularity over time using custom topic visualization*"
- "*[…] spotting changes in real-time is critical, to discover new bugs or issues, for example by identifying a new trending topic or a change in sentiment regarding a tool or a key functionality it provides*"
- "*Comparative view like the 'top topics' is especially important, but should also reflect sentiment […] for example a new tool we have deployed versus the legacy one; or an organic tool versus a third party; or several versions of a tool*"
- "*Compare the use by geography in a similar visualization to the 'top topics' view*"
- "*Get notifications and recommendations for action […] for example, handle the error in our CRM tool, with reference to activities that discuss this error*"
- "*Identify solutions for problems or useful tips to reuse*"

Overall, our eight follow-up interviews revealed many potential merits for the use of the activity stream from an organizational perspective. There were many commonalities among the requirements of the three scenarios, such as segmentation of the data by person attributes (HR attributes, sales-related attributes such as industry or client, general attributes like country or business unit); tracking changes over time with regards to volume of activity or level of sentiment; getting alerts and actionable recommendations in a "push"; or using comparative views (e.g., across peers in sales management or tools in IT). Overall, the high intensity of the stream and its growing usage hold a lot of promise for changing the way LOBs track organizational trends and opinions. Gaining these types of insights from the stream can eliminate the need for solicited feedback via surveys, feedback forms, or other more traditional means.

5 Conclusions and Future Work

Our analysis provides an overview of the way *Streamz* is used within the enterprise. While *Streamz* is still in its early-adoption phase, dozens of employees already use it for a variety of purposes, such as to get a sense of what people say or think, to track the activities of individuals or groups, to search for recent news regarding projects or topics and understand who is involved in them, to search for past discussions, and to generally stay up-to-date.

We observed that stream search is different from traditional enterprise content search in many ways and poses a new complementary paradigm for searching in the enterprise. While content search is used mainly for navigation to pages of internal tools or projects or for finding the main resources related to a topic, stream search is used for other goals, such as getting updated with the recent news or developments, tracking the latest conversations, or understanding the recent buzz. We learned that users typically issue queries as a means to filter the stream rather than to search—as opposed to seeking something specific and known in advance, users often issue queries to render an interesting channel for news discovery.

While *Streamz* originally provided a view for individual employees to help them bisect and consume the activity stream in a personalized way, it was extended in this work with a few initial capabilities that allow better exploration of the stream from an organizational viewpoint. Our follow-up interviews with sales managers, HR, and IT professionals, reveal that this new form of public interaction holds a lot of potential for LOBs who are interested in better understanding what employees do or think with regards to a specific customer, an organizational process, or an internal tool. The UI provided in this work was used to demonstrate an initial set of stream analysis capabilities with an emphasis on the organizational (sociocentric) perspective. More advanced capabilities should allow bisecting the stream by criteria such as industry or person attributes (seniority, salary, etc.), provide more advanced means for tracking trends and sentiment over time, and support active alerts and recommendations for action when anomaly or change occur.

The main limitation of this study is the fact it was conducted in a single enterprise, using a system in its early adoption phase. The results may thus be affected by the characteristics of the specific enterprise and the early adoption. We note, however, that we focused on a rich set of common social media applications that are likely to be relevant in many organizations. We hope future work can further extend and generalize the findings of our research, by inspecting activity streams in other companies and at later stages of adoption.

This paper provides a first glimpse into the use of an activity stream in the enterprise. Currently, the activity stream syndicates activities across social media; however, its potential use is even broader, as it can grow to encompass activities from other organizational systems, such as CRM, software development, or business process management. One of our interviewees even suggested "*I would love to see a Streamz-like unified view of both my social media and email.*" Indeed, email messages can be viewed as another type of a stream activity. With younger individuals joining the workforce, accustomed to using tweets and news feeds as their main means for interaction, an integrated activity stream, equipped with search and analytics capabilities, could potentially replace the traditional mailbox and become the future organizational dashboard for workplace collaboration.

References

1. Ben-Yitzhak, O., Golbandi, N., Har'El, N., Lempel, R., Neumann, A., Ofek-Koifman, S., Sheinwald, D., Shekita, E., Sznajder, B., Yogev, S.: Beyond basic faceted search. In: Proc. WSDM 2008, pp. 33–44 (2008)
2. Bernstein, M.S., Suh, B., Hong, L., Chen, J., Kairam, S., Chi, E.H.: Eddi: interactive topic-based browsing of social status streams. In: Proc. UIST 2010, pp. 303–312 (2010)
3. Boyd, D.: Facebook's privacy trainwreck: exposure, invasion and social convergence. Convergence 14(1), 13–20 (2008)
4. Buffa, M.: Intranet wikis. In: IntraWebs Workshop, WWW 2006 (2006)
5. Busch, M., Gade, K., Larson, B., Lok, P., Luckenbill, S., Lin, J.: Earlybird: real-time search at Twitter. In: Proc. ICDE 2012, pp. 1360–1369 (2012)
6. Byron, L., Wattenberg, M.: Stacked Graphs - Geometry & Aesthetics. IEEE Trans. Vis. and Comp. Graphics 14(6), 1245–1252 (2008)
7. Celi, F., Di Lascio, F.M.L., Magnani, M., Pacelli, B., Rossi, L.: Social network data and practices: the case of Friendfeed. Advances in Social Computing, 346–353 (2010)
8. Daly, E.M., Muller, M., Millen, D.R., Gou, L.: Social lens: personalization around user defined collections for filtering enterprise message streams. In: Proc. ICWSM 2011 (2011)
9. Dork, M., Gruen, D., Williamson, C., Carpendale, S.: A Visual Backchannel for Large-Scale Events. IEEE Trans. Vis. and Comp. Graphics 16(6), 1129–1138 (2010)
10. Efimova, L., Grudin, J.: Crossing boundaries: a case study of employee blogging. In: Proc. HICSS 2007 (2007)
11. Ehrlich, K., Shami, N.S.: Microblogging inside and outside the workplace. In: Proc. ICWSM 2010 (2010)
12. Esparza, S.G., O'Mahony, M.P., Smyth, B.: On the real-time web as a source of recommendation knowledge. In: Proc. RecSys 2010, pp. 305–308 (2010)
13. Farrell, S., Lau, T., Nusser, S., Wilcox, E., Muller, M.: Socially augmenting employee profiles with people-tagging. In: Proc. UIST 2007, USA, pp. 91–100 (2007)
14. Freeman, L.C.: The Development of Social Network Analysis: A study in the Sociology of Science. Empirical Press (2004)
15. Freyne, J., Berkovsky, S., Daly, E.M., Geyer, W.: Social networking feeds: recommending items of interest. In: Proc. RecSys 2010, pp. 277–280 (2010)

16. Gupta, T., Garg, S., Carlsson, N., Mahanti, A., Arlitt, M.: Characterization of Friendfeed: a web based social aggregation service. In: Proc. ICWSM 2009 (2009)
17. Guy, I., Ronen, I., Raviv, A.: Personalized activity streams: sifting through the "river of news". In: Proc. RecSys 2011, pp. 181–188 (2011)
18. Guy, I., Steier, T., Barnea, M., Ronen, I., Daniel, T.: Swimming against the Streamz: search and analytics over the enterprise activity stream. In: Proc. CIKM 2012, pp. 1587–1591 (2012)
19. Guy, I., Ur, S., Ronen, I., Weber, S., Oral, T.: Best faces forward: a large-scale study of people search in the enterprise. In: Proc. CHI 2012, pp. 1775–1784 (2012)
20. Hoadley, C.M., Xu, H., Lee, J.J., Rosson, M.B.: Privacy as information access and illusory control: the case of the Facebook news feed privacy outcry. E-Commerce Research and Applications 9(1), 50–60 (2010)
21. Hong, L., Convertino, G., Suh, B., Chi, E.H., Kairam, S.: FeedWinnower: layering structures over collections of information streams. In: Proc. CHI 2010, pp. 947–950 (2010)
22. Huberman, B., Romero, D., Wu, F.: Social networks that matter: Twitter under the microscope. First Mon. 14(1) (January 2009)
23. IBM Connections – Social Software for Business, http://www.ibm.com/software/lotus/products/connections
24. Kwak, H., Lee, C., Park, H., Moon, S.: What is twitter, a social network or a news media? In: Proc. WWW 2010, pp. 591–600 (2010)
25. Millen, D.R., Feinberg, J., Kerr, B.: Dogear: social bookmarking in the enterprise. In: Proc. CHI 2006, pp. 111–120 (2006)
26. Moran, T.P., Cozzi, A., Farrell, S.P.: Unified activity management: supporting people in e-business. Commun. ACM 48(12), 67–70 (2005)
27. Naveed, N., Gottron, T., Kunegis, J., Alhadi, A.C.: Searching microblogs: coping with sparsity and document quality. In: Proc. CIKM 2011, pp. 183–188 (2011)
28. Ounis, I., Macdonald, C., Lin, J., Soboroff, I.: Overview of the TREC 2011 microblog track. In: Proc. Trec 2012 (2012)
29. Paek, T., Gamon, M., Counts, S., Chickering, D.M., Dhesi, A.: Predicting the importance of newsfeed posts and social network friends. In: Proc. AAAI 2010 (2010)
30. Ronen, I., Shahar, E., Ur, S., Uziel, E., Yogev, S., Zwerdling, N., Carmel, D., Guy, I., Har'el, N., Ofek-Koifman, S.: Social networks and discovery in the enterprise (SaND). In: Proc. SIGIR 2009, p. 836 (2009)
31. Shami, N.S., Muller, M.J., Millen, D.R.: Browse and discover: social file sharing in the enterprise. In: Proc. CSCW 2011, pp. 295–304 (2011)
32. Sun, E., Rosenn, I., Marlow, C., Lento, T.: Gesundheit! Modeling contagion through Facebook News Feed. In: Proc. ICWSM 2009 (2009)
33. Teevan, J., Ramage, D., Morris, M.R.: #TwitterSearch: a comparison of microblog search and web search. In: Proc. WSDM 2011, pp. 35–44 (2011)
34. Zhang, J., Qu, Y., Cody, J., Wu, Y.: A case study of micro-blogging in the enterprise: use, value, and related issues. In: Proc. CHI 2010, pp. 123–132 (2010)
35. Zhao, D., Rosson, M.B.: How and why people Twitter: the role that micro-blogging plays in informal communication at work. In: Proc. GROUP 2009, pp. 243–252 (2009)

Relationship-Based Business Process Crowdsourcing?

Jacki O'Neill and David Martin

Xerox Research Centre Europe, 6 Chemin de Maupertuis, 38240, Meylan, France
{Jacki.Oneill,David.Martin}@xrce.xerox.com

Abstract. New technologies do not always benefit the worker, especially when harnessed by organisations seeking ever cheaper labour. Crowdsourcing is a technology-enabled way of working which offers the potential to bring work to far flung communities. However, it is something of a double-edged sword and there are many socio-technical and ethical challenges. In the micro-task market crowdsourcing platforms tend to be designed largely for the advantage of the organisation requesting work, rather than the worker. This paper contributes to research calling to redress this balance [2, 6]. It describes the findings of an ethnographic study of an outsourced business process – healthcare form digitization – as performed by workers in-office (India) and @Home (USA). It reveals the complexities of the relationships between worker and organisation and argues that designing some aspects of these relationships into crowdsourcing platforms and applications is as beneficial for the organisation as it is for the worker.

Keywords: Crowdsourcing, ethnography, business process outsourcing, relationship-based crowdsourcing.

1 Introduction

New technologies offer new possibilities for working and there are on-going government and private initiatives in most countries throughout the world which aim to harness the power of information and communication technologies (ICTs) for socio-economic development. For example, government initiatives in rural India have led to the use of ICTs to create jobs in rural areas from enabling small producers to take part in large order production[1] to rural Business Process Outsourcing[2] (BPO). However, new applications of ICTs are not always primarily motivated to benefit the worker since work reconfiguration and technology design is often motivated or harnessed by organisations seeking ever cheaper labour pools. Common approaches include a) deskilling, typically dividing complex tasks into smaller parts and using ICTs to assist lower skilled workers to complete these parts; b) offshoring; distributing the work to areas where labour is cheap, or a combination of both a) and b). A poignant example

[1] E.g. http://ropeinternational.com/aboutus.html
[2] E.g. http://desicrew.in/

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 429–446, 2013.

of this is the outsourcing movement, which typically undertakes to complete non-essential or non-central functions for various customer entities at a lower cost. The primary push for outsourcing from the customers perspective is cost reduction through, for example, labour arbitrage i.e. by taking a business function outside of the employer the new workers do not have a right to the same terms and conditions as in-house employees. For the outsourcer, profits are achieved not only through economies of scale, but also through deskilling and offshoring. This is not inherently negative, since for example BPO has had a positive impact on the Indian economy[3], providing much needed employment and propelling infrastructure development. However, it is not a completely rosy picture either, with an frequently undervalued workforce doing low skilled work.

In some domains such as customer care it is questionable whether deskilling is an appropriate strategy, since undoubtedly more professional agents would be better able to support customer needs. However, there is also a large body of work that cannot yet be automated but which in theory could be completed by anyone with a reasonable level of literacy, e.g. data entry of handwritten forms. In this paper we report on a study that was undertaken to see what it would mean to crowdsource this work. Crowdsourcing is a new and growing mode of organizing activities. It has the potential to provide work in developing countries because the workers and work requester (i.e. the enterprise putting out a task to be completed) do not have to be collocated [1]. Thus it is not necessary for the enterprise to set up offices in a particular country, with the attendant cost-benefit analysis, internal and external political choices and so on that that involves. Workers can come from anywhere as long as they have the right skills and a reasonable technical infrastructure. Thus whilst outsourcing has not taken off to the same extent in Africa as in India[4], due to a poorer technology infrastructure and the (comparatively) high costs and difficulties of setting up companies (labour costs, corruption, lack of government incentives, etc.) crowdsourcing could, in theory bring some of this work to Africa[5].

Crowdsourcing is defined as the act of taking a task traditionally performed by an employee or contractor, and outsourcing it to an undefined, generally large group of anonymous people, in the form of an open call[6]. It is frequently used for work which cannot be trivially automated, e.g. requiring semantic knowledge. The term has been widely applied but we focus here on the crowdsourcing of microtasks which are completed for pay and thus fall clearly into the category of work and indeed typically into the category of piecework [2]. Typically a crowdsourcing vendor mediates between the employer and the worker, generally providing a platform through which

---

[3] For example, the Information Technology-Information Technology Enabled Services (IT–ITES) sector (US$100 billion industry) alone has increased its contribution to India's GDP from 1.2% in FY1998 to 7.5% in FY2012 and plays a strong role in generating Employment in India.

[4] Even though South Africa and Ghana might be considered hotspots of outsourcing compared to the rest of Africa, they are nowhere near Indian levels.

[5] Currently the technical infrastructure in much of Africa is poor, however, initiatives such as those to tap into WACS (the West African Cable System) promise to change this

[6] `http://en.wikipedia.org/wiki/Crowdsourcing`

crowdsourcing can be undertaken. Amazon Mechanical Turk[7] (AMT) is probably the best known Crowdsourcing micro-task platform. It enables individuals or organizations to post small, usually low skill, tasks (digitization, translation, search, image labeling, etc.) in large volumes to be taken up by individuals for execution. The workers post back their work for evaluation and get paid on acceptance. AMT has thousands of micro-tasks, which can be executed in seconds or minutes, with payments typically in the order of few cents.

However, like outsourcing, crowdsourcing can be considered to be something of a double-edged sword and there are many socio-technical and ethical challenges. Crowdsourcing offers the potential for providing well needed income for people in developing countries, however employment models tend towards if not quite the idea of getting something for nothing, then certainly something for very little. The appeal to the enterprise is the ability to get work done quickly, without the financial burden of a contracted workforce. Much of the research focuses on (minimum) incentive schemes, how to deal with bad work and spammers and so on. Whilst not denying these are important issues, crowdsourcing tools tend to be somewhat biased towards the needs of the enterprise often to the detriment of the worker [3]. However, as with outsourcing we do not believe that crowdsourcing is necessarily a negative work model, it can just appear that way because, with some notable exceptions [e.g. 1, 4] the tools, technologies and work models employed thus far have been designed to fit a particular organisational perspective. Ethnographic research within domains such as HCI and CSCW is well positioned to provide a critique of this perspective, since it typically reveals the human expertise necessary to carry out even apparently low skilled work [5, 6, 7].

If crowdsourcing is to fulfill its promise of providing employment in poorer communities, companies need to find ways to introduce crowdsourcing into their business processes, thus providing more consistent ongoing work opportunities. This is likely to require considerable process redesign and technology innovation. In this paper we describe an ethnographic study of outsourced form digitisation, which was undertaken as a first step in understanding the possibility/requirements for work and technology (re)design if this work were to be put out to the crowd. Rather than taking the purely organisational perspective of cheaper work done quicker, we want to investigate models where crowdsourcing could be beneficial for both parties - worker and organisation. To this end we consider factors such as the skills and knowledge involved in even this low skilled work and how the nature of the relationship between worker and employer impacts on performance (of both parties!). This study reveals some of the challenges that need to be overcome if such work is to be crowdsourced, given its ongoing nature, strict turnaround times and high quality requirements. In addition it also demonstrates how a more relationship-based approach to crowdsourcing is likely to be as beneficial to the organization requesting the work as it is to the worker. The concept of relationship-based crowdsourcing has been introduced by [3] as part of a call to take the workers perspective in crowdsourcing. Whilst we might hope that all enterprises would aim to treat their workers fairly, in practice this is more likely when

[7] `http://www.mturk.com`

they can see a clear benefit in their day-to-day activities. If crowdsourcing is to be a viable work option to support socio-economic development, it needs to rid itself of its exploitative image. This paper is a first step in showing how, for this type of BPO work, doing so will likely be as beneficial to the organization as to the worker.

2 Crowdsourcing Literature

Much of the crowdsourcing research thus far has focused on AMT and the bias towards the organizational perspective is clear. However recently more research is redressing this balance and we describe work from both perspectives here.

Handling bad data or poor quality work, whether by scammers or genuine workers is a major issue on AMT. Approaches include task design, especially getting the work done iteratively (e.g. find, fix, verify for correcting documents [8]) and/or rating workers through reputation schemes. Whilst it is possible to simply weed out bad work/workers, a more promising approach is to distinguish between scammers and genuine workers and to enable on the job learning so that genuine workers can improve. [10] demonstrated how including initial training sets and gold standard data (i.e. with a known correct output) throughout the task can be used to both judge worker quality (and weed out spammers) and provide feedback enabling genuine workers to improve. Whilst AMT provides means to rate workers' reputations, there is no equivalent means in the tool itself of rating work requesters [3, 11] despite the problem of unscrupulous requesters being well known. To compound this inequality, workers' reputations are harmed if work requesters reject their work. Furthermore AMT deliberately hides the relationship between work requester and worker, often to the detriment of the worker [12]. To begin to address this inequality, Irani and Silberman developed a plugin[8] to AMT which enables crowd workers to rate work requesters [13]. This plugin provides a very useful resource for crowd workers enabling them to identify good and bad requesters on the basis of their collective experience, but it is telling that Amazon have not taken up the clear need for such functionality by embedding it in their tool.

Putting tasks out to the crowd typically involves work redesign and since AMT is a microtask platform, the work needs to be broken down into suitable chunks. An innovative example of this comes from IBM for the OCR correction of scanned documents [14]. They present all the characters which the OCR engine identifies as being the same but with a low confidence rating together rather than having the keyer move from form to form. This improves productivity, removes security concerns (the worker only sees a table of letters with no semantic information) and since it makes the task very low skill, they say it makes it easy to recruit workers on-demand, although this short paper does not document the latter. This research is the closest to our application domain, but not all work is suited to such ultra-micro-tasking, for example the handwritten forms commonplace in our application domain would be hard to split up in this way. Since we are employing people, with all their myriad skills, we believe it makes sense to capitalize on their abilities, rather than to decompose the task into its

[8] http://turkopticon.differenceengines.com/

smallest, lowest skill parts. As Ipeirotis[9] described in his blog, whilst ultra-microtasking can help unskilled new workers, as workers become experienced this embedded workflow can get in the way of them using their knowledge and expertise to produce higher quality work more efficiently. People may also work better, even on simple tasks, when not in complete isolation. [15] found that collaboration amongst crowd members, e.g. asking for help with unknown words during translation, improved quality and promoted learning.

Pay is key concern in crowdsourcing. From the workers side research shows that whilst incentives are important they are not the only factor in determining which work gets done. [11] found that higher pay increases completion rate, time spent on the task and quality but that qualified workers are less affected by pay. In contrast [17] found that pay increased throughput but not quality and [18] found that wage per job was not the sole motivator – rather workers focused on their ability to reach salient targets, e.g. whether there was enough work in this category for them to earn some target amount. Unfortunately much of the research into pay is based on experimental studies rather than real tasks and thus it is hard to really understand how pay impacts performance and indeed take up of jobs in real crowdsourcing work. Certainly from the worker's side, as with employees in any workforce, there is concern about getting fair pay for the work done [3], especially given that very low wages are rife [3, 11, 12]. When we take into account that not only does completion time vary but some tasks never get completed on AMT [19], we can see why some requesters have rejected AMT in favour of other models of working such as ODesk[10] – an online market for contract labour. [19] used ODesk and chose to pay an hourly rate because their job included hard tasks that needed to be completed to high quality and these are the sort of tasks which might likely fail to be completed satisfactorily using AMT. [19] examines fairness in AMT in more detail and discusses why fairness until now has been something of a side issue in crowdsourcing discourses, the design of the platform and the design of crowdsourcing tasks.

Whilst at the moment it might be a choice between AMT and forms of contracted labour, it should be possible to design a different sort of crowdsourcing system which balances the advantages of AMT (rapidity, diversity and access to non-experts) with concerns about ethics [11], completion time and quality[11]. [3] proposes that a more relationship-oriented approach, between requesters and workers, would be beneficial for both – giving requesters higher quality work and workers more fair conditions. As we will describe in the following sections this is particularly important for the sort of work we are hoping to crowdsource. The fieldwork revealed a number of features which would seem to make it more suited to relationship-based crowdsourcing.

Despite its problems, various companies and researchers are already examining crowdsourcing as a means to socio-economic development. Samasource[12] is an

---

9 http://www.behind-the-enemy-lines.com/2012/02/need-for-standardization-in.html
10 https://www.odesk.com/
11 Indeed, researchers have built mobile platforms such as in [4] for more 'ethical' work, but because of the domain constraints our focus is on computer-based crowdsourcing.
12 http://samasource.org/

example of an alternative ethical crowdsourcing company which trains people living in poverty to complete specific tasks for a living wage. Such companies support relationships between work requester and worker – through training, trusted workers and so on. [1] point to a number of issues to be addressed if crowdsoucing is to be extended to low income workers in developing countries including interface design (see [21] for some design solutions), greater regulation of requesters; training; peer-to-peer skills transfer; and predictability of income – which is currently hard to achieve. For our application domain, we would be likely to target the educated underemployed rather than the really low income sectors described in [1, 21] because the nature of the task requires a high level of literacy in English.

3 Method and Study Sites

To understand what it would take to introduce crowdsourcing into a healthcare form digitization workflow, we embarked on an ethnographic study of that work as it is undertaken now in an outsourced environment. Ethnomethodological ethnographies have long been used to understand the nature of the workplace and to conceptualise innovative design solutions (see for example, [22]). Ethnography has proved useful for design because it reveals the underlying skills, knowledge and practices involved in even routine, mundane, low skilled processes. By understanding the work as it is undertaken now and given a particular technology scenario, in this case Crowdsourcing, we can begin to map out what is required to undertake that work in this new setting, as has previously done with for example mobile payments [23]. We believe the ethnographic approach also provides a useful way in to thinking about ethical design, because of its focus on workers skills and the value that they bring to the process.

Approximately five working weeks were spent by one ethnographer (1st author) in the outsourcer's offices in Bangalore and Kochi, India between March and August 2011 undertaking an ethnographic study of the form digitization process for healthcare forms. The entire Indian workflow for three different clients was studied primarily through observation of ongoing work, supplemented by in situ interviewing. The researcher sat with and observed agents doing data entry work at all skill levels, shadowed team leads, supervisors and quality control. In addition a half-day visit was made to the workflow control team, Production Control, who demonstrated to the researcher how they managed their workflows. A further two weeks were spent by a second ethnographer (2nd author) studying home-based workers and support staff in Utah, USA in 2012. The researcher visited agents and supervisors at home and spent some days interviewing and observing the US-based Production Control team. The data collected is qualitative: field notes, audio recordings, photographs and relevant artefacts such as rule books and procedures were collected. It was analysed from an ethnomethodological perspective [24] and used to create in-depth descriptions of the observed work for each client. Ethnomethodology is a non-theoretical analytic orientation uses qualitative data to explicate the endogenous social means and methods by which participants in a setting carry out, organize and reason about their activities.

Presented here are summary descriptions of particular common elements of the work of relevance to transforming those processes for crowdsourcing.

Healthcare forms arrive at dedicated mail rooms in the USA. They are scanned and if possible OCR'd. The scanned forms are routed to the relevant onshore and/or offshore workforces (depending on client agreements). The following steps are undertaken for all clients: 1) Data entry: keying forms not suitable for OCR by hand or checking OCR errors; 2) Verification: a second agent enters the data and is prompted if there are discrepancies; 3) Review: experienced agents check specific fields. There are a number of other activities that are carried out to ensure the smooth running of this workflow, including 1) quality audits and six sigma projects; 2) supervision and management, including shift organisation, reporting and floor-walking; 3) monitoring and managing the flow of forms through the process by production control.

4 Findings

As the work is already outsourced one might think that the challenges of distribution have been solved and to an extent this is true in the sequential steps of the workflow through which each claim passes. However, our findings show that even low skill data entry for non-OCR forms poses a number of challenges for crowdsourcing. We group these into four themes which emerged from the data: workplace ecology, skills and knowledge for data entry, making the workflow work and collaboration. For each theme we discuss the features in each setting (in office and @home) and then describe the implications for crowdsourcing. In the discussion we focus on the points which lead us to propose relationship-based crowdsourcing as an appropriate solution. Before we turn to the themes however we say a few words on the nature of this sort of BPO work and the workplace in which it is undertaken.

This work might be characterized as 'white collar factory work' and this idea deserves some unpacking. In the earlier 20th Century clerical work was considered a higher status work, offices were the places of work where the quality of environment was better than a factory, dress was more formal, and pay was generally higher. In the latter 20th Century, however, in a trend that continues, manufacturing principles were applied to administrative business processes. This involved an increasing decomposition, separation and standardization of administrative tasks within business workflows. This activity particularly impacted on the most basic tasks – such as data entry and digitization – that apparently needed little professional knowledge and skills. These tasks were hived off to be completed as low-pay, low-status, high volume piece work, especially when they became outsourced. Outsourcers, in wishing to keep their costs down, typically spend little on facilities, with the technical infrastructure being most important.

This is borne out by our experiences; the offices of the workers, despite being situated in fancily named 'tech parks' are typically basically fitted out. The work is mainly piece rate with only certain supervisor duties or management paid by the hour. Consistent high performance can earn employees a bit more than minimum wage, while poor performance over a period of months will earn employees the sack.

However, compared to similar call centre environments, the workers are at least in control of their own pace of work. Indeed, some employees attain a mix of skills and knowledge that allows them to consistently perform very well, while not placing a stressful burden upon themselves. They attain high speeds, high concentration and have particular types of sensory-motor skills (e.g. being able to notice at-a-glance that something looks either correct or problematic) while also being able to attain the required knowledge. These workers are prized and often promoted. This understanding begins to give the lie to the idea that this is wholly routine work that just about any motivated person could do well. Particular skills are involved that not everybody possesses. Retention of good staff is the challenge and goal in this type of industry, because they are not always so easy to find, and recruitment and training costs are relatively high. It is important that the employer has confidence in their workforce delivering consistently. For the employee, of key importance is the ease with which they can carry out the work, the stress levels they deal with, the stability of work volume and the predictability (and control of) their wage.

4.1 Workplace Ecology

The workplace ecology for @home workers in the US and in-house workers in India is unsurprisingly quite different. Each provides various challenges and benefits for the configuration and management of the work; technical, legal, organizational and social. In India the outsourcer's offices in both locations span several floors of office blocks in technical parks. Employees work on data entry for single clients and are grouped together by client in particular seating areas or access-controlled units. The essence of the work site is a highly controlled environment, whereby security and surveillance prevails. Security is ensured technically (through systems) and through the ability to monitor staff activities through the shared physical space, which affords visual and auditory monitoring. The actual environment is standard office space with some open plan and divided areas. By contrast for people working at home in the US, their office space is naturally in their home – with each being different. Their basic requirement is to have a separate office/study space. This might be a dedicated room or a space within a room that is dedicated for work. It must be private (for the period they are working), i.e. if others are in the house with them during their working hours they must not be sharing the work space. Being home offices, the design and arrangement is up to the worker, apart from the technologies supplied by the company. They have access to the rest of their house and the home comforts that provides (for example, seasoned workers can do much of the job while listening to the radio or TV). The home environment affords greater freedom than the in-office environment. Although agents work can be monitored through the system, e.g. the supervisor can view the agents desktop at any time, there is not the co-present supervision at play in the office.

Data Security in-Office. Since health care forms contain personal information including social security numbers, names and addresses, data security is governed by

US laws, in particular HIPPA compliance. This strictly controls who may access the data and protects against unauthorized distribution and use. Data security is currently enforced through physical, contractual, social and technical means. Physically, access to the workspace is restricted, with passes required to enter the technical park and offices and security guards at various entrance points throughout. Employees also have to leave their bags and phones in lockers during the shift and pens and paper are not allowed in the office. Contractually, employees agree to the various company regulations and non-compliance has consequences including dismissal. Socially security is enforced through supervision with team leads walking the floor and answering queries, whilst keeping an eye on the workers. Workers are trained in HIPPA compliance. In addition there are a variety of technical solutions, with the systems and workflows designed to ensure maximum security. For example, using thin clients which do not store information and have no USB ports. Plus data is stored in the US and is pulled in in batches for processing. Each batch only resides locally during processing. Agents cannot access client systems and can only see the current form.

Data Security @Home. The home workers' environment is less secure by nature, because there is no co-present supervision. However, their environment must meet certain criteria (access to broadband, a separate space to work and a conducive family environment). The workers are supplied with a secure terminal which they route through their internet connection in order to connect to the company's network and work in secure thin client mode. All staff members working from home have completed all necessary training, including HIPPA compliance, and are generally experienced and trusted members of staff. Thus, data security is currently enforced primarily through contractual and technical means. Physical and social means still play a part, but cannot be so tightly controlled and monitored, since although they have to have a separate working area, there is no way of ensuring no one else is present. Ultimately there is a reliance on trust. Supervisors cannot walk the floor and have an overview of what everyone is doing at any one time and cannot police the use of cameras and so on, but they can view an operator's desktop at any time. Working at home is an earned privilege. In order to be selected, employees need to be experienced and high performing, and informal assessments are made as to whether they are good candidates for the remote and autonomous environment. If there are any concerns about their work they can be brought back into the office environment.

Implications for Crowdsourcing. The main implication for Crowdsourcing comes from the distribution of the workers from controlled office or home environments into uncontrolled home environments or unsecured Internet cafes and their lack of a contractual relationship with the company. The control the outsourcer can exercise over the people doing the work is necessarily reduced. In effect then, security can no longer be enforced by contractual, physical and social means and solutions to data security will have to be wholly technical. One solution is to take the IBM approach [14] and split the data such that semantic information cannot be derived from the task by the worker. However, this is not appropriate for all tasks, especially for handwritten forms which cannot undergo OCR. In this case, an alternative, and conceptually simple, solution is to separate out the parts of the form with sensitive data from those without. If a hybrid workforce were engaged i.e. a mixture of contracted and crowd-sourced labour, the non-sensitive parts of the form could be directed to the crowd-sourced workforce, whilst the sensitive parts could be done in-house.

4.2 Skills and Knowledge for Data Entry

In-Office. Agents have a basic level of education (graduate or undergraduate), good English language skills and typing speed. The work is known as 'key what you see' and is considered low skill; nonetheless the learning curve of a new entrant is around seven weeks. This is because in reality data entry is not simply 'key what you see' rather agents must interpret what they see according to an extensive rule set. To illustrate, the most straightforward data entry task is a standard form for claiming medical insurance (a HCFA). The name field alone has around 13 rules for how the name should be entered (for example decided which is the first name surname, middle name and so on) and it is just one of 33 fields. Each client has a variety of different forms and other documents: agents need to learn how the rules apply for all the different form types. Furthermore, task complexity is situational. It differs between and within form types. To illustrate, although data entry for correspondence is only four fields compared to 33 for HCFAs, correspondence can take considerably more time as the information may be anywhere or nowhere on the form. Identifying that information is not present often takes longer than finding and entering information because of the need to double or triple check. Agents are paid per keystroke or per form (depending on job type) with quality taken into account, thus the speed with which they can work is of immediate concern to them.

Complexity also differs across individual claims: 1) Handwriting or poorly printed forms can be difficult to read. 2) If forms do not fit the criteria for data entry because of poor scanning or because they do not comply with the rules (e.g. two patients named on one form) they need to be rejected and this takes more time because it requires double or triple checking. 3) Non-standard means non-standard; a piece of correspondence may have a cover sheet with all the required information on it or the information may not be found in the document at all. Within form complexity can only be determined on a document by document basis and would not be easy to predict in advance.

@Home. The at home workers are trained in the office and the best workers (fastest, highest quality, most reliable) are permitted to work from home if they so wish. Interestingly whilst the skills and knowledge requirements apply equally to the agents working at home, the difference in the nature of the two workforces (Indian and US) was obvious during our observations. The Indian workforce was a more transient one, with many agents leaving and needing to be replaced by newcomers. Whilst some agents had been working there for a couple of years, most were considerably newer, with all teams having a number of new starters on board. In comparison, the at home workforce consisted of the best agents (as only they are allowed to work from home) and throughput was minimal, with agents typically having been employed with the company for a number of years (some more than 15). Thus the at home workforce, as a whole, was highly experienced and they worked with apparent ease interpreting data and spotting and correcting errors with little hesitation. Indeed it was common for agents to listen to the TV whilst working as they had become so skilled at data entry.

It is important to note we are not implying some fundamental difference in the workforces skills, indeed the more skilled Indian agents worked just as adeptly although they did not get the opportunity to listen to anything. Instead the difference is one of stability of workforce and the benefits which are accrued from that, i.e. the @Home workforce had on average more knowledgeable and skillful workers due to their greater experience. Working from home confers a number of advantages on the employees who choose it. The primary one being flexibility of working hours – workers choose at what time to work their hours each day, enabling them to fit their work around their other commitments (the vast majority of the workforce are stay at home mums). Other advantages also come from reduced supervision, a more relaxed and comfortable environment and regular and predictable income. For many the work is attractive because it provides what might be considered to be an acceptable compromise; it's not the most interesting job and many workers have the ability to do higher skilled work, but it's 'local, it fits into their lives and so is the preferred option.

Implications for Crowdsourcing. Whilst it is clear that an office environment is not essential for this work, it is also clear that both agents and employer benefit from having a well-trained experienced workforce, particularly given the strict quality requirements in the SLA between the client and the outsourcer. For crowdsourcing, we need to think of how to manage the advantage that experience bestows, plus how the learning curve might be accommodated. Approaches could include 1) reducing the required learning; e.g. splitting the form in to sections, reducing the number of rules agents must learn, contextual presentation of rules and embedding training within the dataset [10], and 2) encouraging workers to become skilled in particular areas; whilst specialization does not have to be requester specific and indeed one of the founding concepts of crowdsourcing is that it would not be, for the work requester, there is likely to be benefit from being the 'first choice' for the best workers. Being the favored work requester for the best workers, is likely to be a major drive towards relationship-based crowdsourcing (see Discussion).

The fieldwork also raises the question of how to determine incentives to ensure that all work is completed, given its situational complexity. Pay scales are currently determined by form type, but social and organisational pressures encourage agents to complete the batches they pull, whatever their complexity i.e. they are held accountable by supervisors. If crowdsourced, what's to stop agents from rejecting difficult work? The outsourcer is unlikely to move to hourly wages for crowdsourcing, given that they already use piece rate for their contracted employees. We therefore must assume such rejections will occur and identify and handle them automatically whether through dynamic incentive and reputation schemes or hybrid models of in-house and crowd workers.

It is also clear that there could be potential benefits from being a crowdsourced worker (over an in office worker) as, given a fair rate of pay, the benefits which stem from working conditions at home would carry over. At the moment, homeworking is only available within the US, largely because of data security concerns, but if a wholly technical solution to data security could be found then this work could be crowdsourced and thus done from home anywhere in the world with a good internet connection. We are not underestimating the restrictiveness of this latter condition, but as connections improve the work can be done further afield. For such work to be possible in internet cafes, it would need to be adequately compensated to pay for the use of the infrastructure and the facilities [25].

4.3 Making the Workflow Work

In-Office. The Service Level Agreement (SLA) is the contract between the outsourcer and client, covering data security, Turnaround Time (TAT) and quality. Whilst pay is a strong motivating factor for agents, it is not enough on its own to ensure the SLA is met. Team and quality leads put in extra work to make the agents accountable for their performance. To illustrate, for one client the data entry of medical records, must be completed by 8am. A key concern of the 6am shift is to complete all remaining records before the deadline. The team leaders communicate the pressures of the queue to the agents ('agents key fast' 'everyone on medical records'). In addition once the deadline has passed the team leader calls up three or four agents at a time to discuss (rather publically) their performance. For example, "Krishna, you only did 16 medical records in an hour. What was that about?" "Ok the target of 50 is not attainable but I would expect you to be doing at least 30. You need to improve." Thus the agents are called upon to account for their performance and team leads work to make the targets achievable for *their* team. Rather than telling someone only doing 16 medical records they should be doing 50, which might seem unachievable and be demotivating, they set a more realistic target. Agents are similarly made accountable when they reject forms, as team leads monitor rejections and question agents. If a batch is rejected a few times they assign it to someone whom they insist must complete it. In this way a balance is achieved between the agents' desire to do the easy work quickly and the requirement to get all the work completed in a timely manner.

@Home. Home work is governed by the same criteria as the in-office work. However, more flexibility is built into scheduling and in general supervision is more light touch. The volumes of work were fairly dependable, however scheduling was still decided on a daily basis according to the volumes of work. The managers would do a daily calculation based on volume and staffing levels and determine a number of hours per employee, which they could complete according to their preferences (and constraints) within the day, e.g. they could work at whatever times they wanted, splitting shifts as they wished, as long as they completed their hours. Generally, the highest volumes came in on Monday and Friday with the lowest volume on Wednesday. This meant that to some extent they could organize their home life according to these trends. Supervisors did intervene if they were concerned about worker productivity on any given day (by email, phone, pager or IMS), or there were particular pressures or requirements during a given day. The vast majority of home workers are high performers so while assessments and training and motivation is dispensed according to employee needs it is not a particularly regular or notable feature of the working day.

Production Control. On the 'shop floor' there is a concern for making sure the volume of jobs is adequately managed as we noted in the sections above, for example by making sure someone takes up a previously rejected batch, by pushing workers to up their speed, or getting homeworkers to change their shift hours. Production control, on the other hand, is the function specifically responsible for ensuring that the workflow flows smoothly and that jobs and batches do not get held up such that they miss their deadlines. To do this they have various tools at their disposal. A workflow management tool moves batches between several dedicated applications (on dedicated servers), including scanning and uploading, OCR and a tool for handling all the (human and machine) steps concerned with data processing. Given the shift patterns and productivity across locations, one of the tasks of production control is to see whether work should be reassigned. They assess how the processing work is unfolding across different locations, technical difficulties, skill coverage and so forth. They also check that batches are complete as they move between applications and that files or cases (or indeed batches) are not lost or held up. Hold ups can derive from technical problems, e.g. OCR application crashes or human factors, e.g. has someone on the shop floor has checked and reassigned documents in the rejected queue. The production line is constantly rolling and it constant monitoring is required to ensure nothing is missed.

Finally, for the various data entry steps (keying, verification and review) the workflow needs to be dynamically managed such that each claim passes through each step in a timely manner. Agents are assigned prioritized task permission lists, i.e. each employee has a list of tasks that they are qualified to undertake and these are prioritized. For example, an agent with OCR correction as priority 1, will be assigned an OCR correction task as their next task each time one becomes available. If no OCR tasks are available, the task rated as priority 2 *for that agent* will be assigned. Given an the unfolding work and a given set of employees, production control decide whether to manually reorder these lists and re-prioritise and re-assign tasks to ensure their timely completion. Thus the work of production control involves both technical and organizational interventions and is crucial to ensuring that the workflow unfolds in a timely and complete manner.

Implications for Crowdsourcing. The current model of work is a push model with a) the work being assigned dynamically (by the workflow tool and production control) according to agents' skill set and queue and b) various social (and financial) pressures. In contrast crowdsourcing involves a pull model: work is self-selected by the agents who will not have the same accountability. The importance of meeting deadlines and managing the pace of the workflow for the outsourcer are clear.

Any crowdsourcing system needs to be designed to ensure that the work is completed in a timely manner to good quality, for which coordination of the workflow will be key. A major issue is that, without doubt, including crowdsourcing in the workflow will increase its complexity. There will be more places for faults to occur and documents get lost or held up. What will happen once a batch has been sent to a crowdworker? How will you know if a crash has occurred at their local PC or whether they have stopped working on the batch? How and when can you pull a batch from them and inject it back into another part of the organization/workflow? These are potentially complex issues both technically and organizationally, and there are clearly implications in terms of how relationships function and are managed with the crowd. Certainly as much of the workflow management as possible would need to be automated to deal with this increased complexity. However, it seems likely that even with the best automated tools a greater burden will be placed on production control.

In addition, it is interesting to note that even in this highly controlled piece rate environment, financial incentive is not considered enough to ensure that targets are met and all sorts of social factors come into play in determining how agents perform. It would be foolish to think that crowdsourced workers will have any less complex motivations, as indeed the research on incentive schemes has hinted at [3, 11, 17, 18]. Whilst pay is undoubtedly highly important, other factors should not be neglected (such as notions of respect, trust, fairness).

4.4 Collaborative Working

Collaboration has largely been designed out of the workflow. Claims progress through workflow steps from agent to agent and country to country automatically as each prior step is completed.

In-Office. However, in the office the work is collaborative at the claim level. That is, the routine troubles encountered in data entry are solved with colleagues or floorwalkers and it is not uncommon to see a group of two or three people around a screen discussing an issue. Typical issues include deciphering handwriting or determining which rule applies to this circumstance. All sorts of phenomena may be found on the forms (arrows showing a name has been entered the wrong way round, handwritten clarifications, etc.) not all of which are described in the rule set. Team leads walk the floor answering such queries, however they are not always immediately available and since time is money for the agents they may turn to their colleagues for help. Where the issue is deciphering handwriting, agents will typically ask their neighbour first, only turning to the team lead if this fails. For questions of rules the team lead, or other 'expert' users are the first port of call. Such troubles rarely take long to solve, but it is undoubted that this collaboration improves both speed and quality. In addition, newcomers, and agents who have fallen below a quality threshold, have their inputs checked by a supervisor before saving them in the system. This supports their learning whilst maintaining the high quality scores of the whole team.

@Home. A clear difference between at home work and in-office is that collaboration is minimized at home – there are few opportunities for at home workers to collaborate amongst themselves, share knowledge and experience. Communication is between the employees and the supervisor either in the form of 1-1 or global communication between the supervisor and the team. It is largely electronic, they use IM and email, supplemented by mobile phones and pagers, in some lines of business. The phone generally used as a means of stepping up, i.e. due to the failure of an electronic channel to achieve the desired aim due to lack of response or complexity of the communication.

Implications for Crowdsourcing. The @Home scenario might lead us to think that collaboration is not important for the work and therefore not of importance for crowdsourcing. However, it should be remembered that the homeworkers are trained in-office – so undertake their learning in the company of peers – and are the highest skilled workers. It is more likely that the by and large the crowdworkers will more closely resemble new in-office workers than the @Home workers. We might therefore want to think about enabling collaborative set ups amongst crowd members [15] enabling them to help one another, or to have sub-crowds with particular skills, such as handwriting deciphering, to which data fragments might be sent.

5 Discussion

There are a number of ethical issues which have been raised around crowdsourcing in the relatively low skill, micro-task domain, including low wages and the bias towards the work requesters' needs. However, crowdsourcing itself is not inherently unfair. Rather these issues are an artefact of AMT (and other platform) design and that much of the research so far has been driven by work requesters concerns, where the requirement to drive down costs can too easily override all other interests especially ones which are hard to operationalize such as fairness [20]. We believe that the particular constraints revealed in our domain through the ethnographic study, provide a starting point for understanding how addressing these concerns not only benefits the worker but also the organisation who is requesting the work. The idea of relationship-based crowdsourcing, described in the literature, came out of research examining the crowd workers perspective when undertaking traditional AMT tasks. When thinking about crowdsourcing this type of on-going BPO work, it can be seen to make sense from the organisational perspective as well. The fieldwork has revealed that even for this low skill piecework, it is the subtle and complex relationship between the organisation and worker that plays a major role in getting the work done such that it can meet the strict quality and turn-around time requirements.

The fieldwork highlighted a range of considerations and challenges for crowdsourcing. It is not the aim of this paper to address them all[13], although we hope that

[13] A discussion of some of them, in particular form decomposition (to address data security and other issues) and the extreme distribution of the workflow can be found in [26].

raising them will provide a resource for the designers of crowdsourcing platforms and plug-ins. Instead, we focus the discussion on the findings which demonstrate the subtleties of the employer-employee relationship in the work as done now and those that lead us as a consequence to propose relationship-based crowdsourcing as a model for this work. A major argument for relationship-based crowdsourcing is that both agents and employer benefit from a well-trained, experienced workforce (for both speed and quality). Even work, which on first glance, is low skill and indeed is even known as 'key what you see' actually has a considerable learning curve. Whilst we can certainly think of techniques to reduce this learning curve (specialisation, embedded contextual rules, training data and feedback embedded in batches), the same is likely to hold true for crowdsourcing. That is, if the organisation can attract a (relatively) well skilled crowd who spend time becoming experts in their tasks it is likely to benefit both. One way to attract such a crowd, is to become (one of) the best work requesters. Perhaps equally important is that our fieldwork reveals that even in this low skill piecework environment, pay whilst important is not enough on its own to get the work done in time to the right standards. Rather a variety of social factors come into play. There is no evidence to suggest that pay will be the only driving factor in crowdsourcing either. Indeed the mixed findings of the incentive schemes studies [3, 11, 17, 18] and the very call for relationship-based crowdsourcing [3] suggests that it would be foolish to think that it would. So if pay alone is not an adequate driver, what is? We propose that ideas of trust and fairness in the relationship between the requester and worker are likely to play an equally important part. Nonetheless accountability of workers will be reduced and any crowdsourcing system needs to include tools and technologies which enable the smooth running of the workflow with minimum human intervention, otherwise the role of production control will become overly burdensome and problematic issues will be missed. Finally, whilst data security might seem like an obviously technical question, and indeed many of the solutions in place now are technical, it too is social. As seen in the homeworking situation, acceptable security is actually reached in the relationship between the worker and the organisation. It is contractual, but it goes beyond that, stemming from the homeworkers being valued and trusted employees. However, when it comes to crowdsourcing, even with relationship-based crowdsourcing, the relationship will not be of the type or strength to enable trust on this scale, and thus data security solutions will have to be technical.

What then do we mean by relationship-based crowdsourcing in this situation? As suggested in the literature [3, 20], the idea is to create a fair and balanced crowdsourcing relationship between the organisation requesting the work and the worker doing work. Subtle notions of trust and fairness come into play. It naturally includes factors such as decent pay for the work done, paying on time, fair and transparent accepting or rejecting of work, embedded training and feedback, ways of rating work requester as well as the worker and so on. As part of the drive for higher quality on both sides of the relationship, it may also include supporting relationships between crowdworkers, enabling them to learn off one another [1] and even collaborate.

It is certainly not about trying to recreate employee-employer relations, just without the same contract and benefits. We do believe that crowdsourcing as a way of working and getting work done can confer benefits on both workers and requesters.

Our fieldwork showed that working from home, at least in the US, offers certain comforts and whilst some of these will be dependent on socio-economic status, others, such as not having to migrate or travel to work, to manage your work around your other commitments and to fit work into 'spare' hours will be more universal. There are certainly many people worldwide who would see benefits form working for home. Even working from internet cafes will have some benefits (e.g. less direct surveillance) but additional economic costs. Although not suggesting that crowdsourcing is a panacea to underemployment and socio-economic development, we do believe that it offers real opportunities to bring work to various underserved populations. As crowdsourcing opportunities grow, workers should be able to pick and choose the work which is of interest to them, capitalizing on the ability to work for many different organisations depending on what work is available. Additionally the work described here has advantages over other more typically crowdsourced work, since it is a large-scale business process it can offer more predictable income to workers. For the organisations requesting work to be done, even with fair wages, they will gain cost savings from massively reduced infrastructure costs, plus the ability to easily ramp up and ramp down the workforce and if taking the relationship-based crowdsourcing approach, greater chance of accessing the best workers in the crowd, for that type of work (as with the homeworkers).

However, the current crowdsourcing model is too biased towards the organisations requirements and the worst of crowdsourcing – piecework with unfairly low remuneration - does not fit with the benefits which accrue from experience: better performance, quality and loyalty from the best workers. If crowdsourcing for this type of BPO work is to be sustainable over time, then a different approach needs to be taken. Whilst relationship-based crowdsourcing may be a somewhat optional model for some types of tasks (e.g. one offs such as image labeling) for this work, because of all the factors discussed (TAT and quality and on-going nature of work; skills and knowledge and even security) it would seem to be the most logical choice for both organisation and worker. The question remains of how to organize and implement it? Just as the organizations' perspective has been designed into current platforms such as AMT, the platforms and interfaces to support relationship-based crowdsourcing will need to embody new more equitable labour market models, which is a major research challenge for the future.

References

1. Thies, W., Ratan, A., Davis, J.: Paid crowdsourcig as a vehicle for global development. In: CHI Workshop on Crowdsourcing and Human Computation (2011)
2. Felsteiner, A.: Working the Crowd. Employment and Labor Law in the Crowdsourcing Industry. Berkeley Journal of Employment & Labor Law 32(1)
3. Silberman, M.S., Irani, L., Ross, J.: Ethics and tactics of professional crowdwork. XRDS 7(2), 39–43 (2010)
4. Eagle, N.: txteagle: Mobile Crowdsourcing. In: Aykin, N. (ed.) IDGD 2009. LNCS, vol. 5623, pp. 447–456. Springer, Heidelberg (2009)
5. Sachs, P.: Transforming Work: Collaboration, Learning and Design. Communications of the ACM 38(9), 36–44 (1995)

6. Suchman, L.: Making Work Visible. Communications of the ACM 38(9), 56 (1995)
7. Hartswood, M., Procter, R., Rouncefield, M., Slack, R.: Making a Case in Medical Work: Implications for the Electronic Medical Record. CSCW 12(3), 241–266 (2003)
8. Kochhar, S., Mazzochi, S., Paritosh, P.: The Anatomy of a Large-Scale Human Computation Engine. In: HCOMP 2010, pp. 10–17. ACM, New York (2010)
9. Bernstein, M., Little, G., Millar, R., Hartmann, B., Ackerman, M., Karger, D., Crowell, D., Panovich, K.: A word processor with a crowd inside. In: UIST 2010, pp. 313–322. ACM (2010)
10. Le, J., Edmonds, A., Hester, V., Biewald, L.: Ensuring quality in crowdsourced search relevance evaluation: The effects of training question distribution. In: SIGIR 2010, pp. 17–20 (2010)
11. Adda, G., Sagot, B., Fort, K., Mariani, J.: Crowdsourcing for Language Resource Development. In: 5th Language and Technology Conference, Poland (2011)
12. Bederson, B.B., Quinn, A.J.: Web workers unite! addressing challenges of online laborers. In: Proceedings of the 2011 Annual Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA 2011), pp. 97–106. ACM, New York (2011)
13. Irani, L., Silberman, M.S.: Turkopticon: Interrupting Worker Invisibility in Amazon Mechanical Turk. To appear in CHI 2013 (2013)
14. Karnin, E.D., Walach, E., Drory, T.: Crowdsourcing in the document processing practice. In: Daniel, F., Facca, F.M. (eds.) ICWE 2010. LNCS, vol. 6385, pp. 408–411. Springer, Heidelberg (2010)
15. Rzeszotarski, J.: Worker Collaboration in Crowdsourcing Markets. In: ACM SIGCHI 2011 Workshop on Crowdsourcing and Human Computation (2011)
16. Kazai, G.: An Exploration of the Influence that Task Parameters have on the Performance of Crowds. In: CrowdConf 2010, San Francisco, USA (2010)
17. Mason, W., Watts, D.J.: Financial Incentives and the "Performance of Crowds". In: HCOMP 2009. ACM, Paris (2009)
18. Horton, J.J., Chilton, L.B.: The labor economics of paid crowdsourcing. In: EC 2010, pp. 209–218. ACM, New York (2010)
19. Huang, E., Zhang, H., Parkes, D.C., Gajos, K.Z., Chen, Y.: Toward automatic task design: a progress report. In: HCOMP 2010, pp. 77–85. ACM, New York (2010)
20. Silberman, M.S.: What's fair? Rational action and its residuals in an electronic market. Unpublished manuscript (2010), `http://www.scribd.com/doc/86592724/Whats-Fair`
21. Khanna, S., Ratan, A., Davis, J., Thies, W.: Evaluating and Improving the Usability of Mechanical Turk for Low-Income Workers in India. In: ACM DEV 2010 (2010)
22. Bentley, R., Hughes, J.A., Randall, D., Rodden, T., Sawyer, P., Shapiro, D., Sommerville, I.: Ethnographically-informed systems design for air traffic control. In: CSCW 1992, pp. 123–129 (1992)
23. Kumar, D., Martin, D., O'Neill, J.: The times they are a-changin: mobile payments in india. In: CHI 2011, pp. 1413–1422 (2011)
24. Garfinkel, H.: Studies in ethnomethodology. Prentice-Hall, Englewood Cliffs (1967)
25. Wyche, S.P., Schoenebeck, S.Y., Forte, A. "Facebook is a Luxury": An Exploratory Study of Social Media Use in Rural Kenya. In: CSCW 2013 (in press, 2013)
26. O'Neill, J., Roy, S., Grasso, A., Martin, D.: Form Digitization in BPO: From outsourcing to crowdsourcing? To appear in CHI 2013 (2013)

User Experience and Learning Experience in Online HCI Courses

Saila Ovaska

School of Information Sciences, University of Tampere
Kanslerinrinne 1, 33014 Tampere, Finland
Saila.Ovaska@uta.fi

Abstract. Several Massive Open Online Courses (MOOCs) have become available also in the field of Human-Computer Interaction (HCI), in addition to other fields of study. How do they operate, and what do they teach? We compare two HCI MOOCs with a locally composed small scale online HCI course that was built and launched in 2012, and discuss issues in pedagogy, user experience and learner experience in online HCI education.

Keywords: HCI Education, online education, MOOC, User Experience.

1 Introduction

A Massive Open Online Course (MOOC) is a Virtual Learning Environment (VLE) where anybody interested in the topic of the course can sign up and participate remotely. Distance education without personal face-to-face interaction between educators and students is not a novel idea—online education has existed in various forms for decades [2, 19, 20], and Learning Management Systems (LMS) such as Blackboard and Moodle have become ubiquitous [16].

According to Daniel [6], the first online “MOOC” was arranged in 2008, and, contrary to most current MOOC instances, it was based on philosophy of connectivism and networking, life-long learning and distributed content [5]. Daniel calls it “cMOOC”: it differs quite a lot in pedagogy from more recent “xMOOC”s that usually evolve around a more standard course structure with lectures, assignments and other content available on the course site or in open streaming video platforms.

What makes MOOCs special is their open nature: they do not link directly to university degree studies but are open and free to all interested. There is quite a lot of variation in how they are organized, but many are arranged by a special platform and service provider (such as Coursera, edX, Khan Academy, and UdaCity, to mention some) that collaborates with educators who provide the content material. Because of the novelty of the idea, the policies vary a great deal from one course to the next, including questions about practical things such as availability of certificates and possible fees of accreditation, ownership and copyright of the content materials, and relationship with established course modules at universities. Still, MOOCs seem to bring new opportunities for learning to many, in particular to those who live in the

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 447–454, 2013.

developing world [15] or cannot afford the high tuition fees of on-campus education [6]. All age groups can take part, even in rural areas with shortage of competent teachers [6].

A pertinent question posed by educators is whether MOOCs will eventually change the way we teach [16, 22]. Furthermore, students are increasingly likely to request institutions to acknowledge their prior learning obtained through MOOCs.

Reports about early experiences with MOOCs in various areas are abounding, at least from the educator perspective [4, 6, 10, 11, 15, 16]. The reported first experiences with MOOCs are varied. Drop-outs are common. Only a fraction of those who initially register finish all required parts of the course. On the other hand, much of the feedback is positive and encouraging, so that the courses will be improved on and run a second time. However, many experience reports point out the need for pedagogy and learning expertise in the design of MOOCs.

2 Background and Motivation for Our Work: HCI Education

In this paper we discuss three courses in the field of HCI. The Stanford HCI Course, arranged by Scott Klemmer [13] in Coursera, launched its third instance in April, 2013. The first two times when it was given reached more than 100,000 registrations [22]. Another HCI Course was opened to the public by Alan Dix [8] in early 2013. Both of these are open access MOOCs but with a bit different focus and methods.

We created an online HCI course (nicknamed "CUX", "Course on User Experience") in-house during 2012. CUX is not open to everybody and by no means "massive". It is designed for the students at our university who start their degree studies for MSc in Human-Technology Interaction. The MSc program is targeted to students who have a bachelor's degree in a suitable field, such as Computer Science. When starting in the program, some of the new students lack formal education in the basics of HCI, at least to the extent already covered in our own BSc program.

The main motivation for arranging an online HCI course for the students is to give them a compact understanding of the basics that are a prerequisite for the courses in the MSc program. The yearly intake in the MSc program is less than 20 students, and only a handful of them are required to start their studies on the introductory CUX course.

The MSc program started in 2001, and during the years we have tried out several pedagogical solutions for the introductory classes, ranging from participation in ordinary teacher-led class meetings to blended learning and self-study arranged with a Moodle LMS as a central repository of readings, but without any guidance and assignments the online studies solution proved inefficient. Structured assignments with deadlines and personalized guidance and feedback are needed for learning [20].

Our focus in this paper is pragmatic. While CUX development coincided with the advent of the MOOCs in HCI, now that they exist, could parts of these courses be utilized locally to support and perhaps partially even substitute locally developed content? What differences are there in these three courses? We will next turn to look at the courses in more depth.

3 The HCI Courses as Learning Experiences

Many studies (such as [21, 23]) focus on user experience in Virtual Learning Environments, but a closer inspection of these studies reveals that the VLEs in question are 3D virtual worlds such as SecondLife and not MOOC platforms. The importance of good learning experience in all web-based learning is recognized [1, 12, 18]. All activity within the VLE is carried out through the interface, making usability and user experience integral parts of the learning experience, and affecting the learning outcomes [9]. The educational setting will need to encompass both pedagogy and technology [7]. Students need to be motivated to study the material, and the learning tasks and support for them need to be designed so that they enforce effective learning [12].

The two MOOCs and also CUX can be described as Learning Management Systems where the lecture presentations are arranged into study units together with topical assignments and quizzes. Table 1 provides short summaries of the HCI courses in March 2013. The information of the two MOOCs is based on their web pages [8, 13].

Table 1. Course Comparison Table

	Stanford HCI Course [13]	HCICourse [8]	CUX
Aimed at	Beginners	Beginners	Beginners
Goal	Build design skills	Deliver free and open access info in HCI	Introduce HCI principles and key concepts
Units/ Weeks/ Main topics	1: Introduction 2: Needfinding 3: Rapid prototyping 4: Direct manipulation 5: Heuristic evaluation 6: Representations 7: Visual design 8: Information design 9: Designing experiments 10: Running experiments	1: Introduction 2: Interaction design 3: Human perception and cognition 4: Emotion and experience 5: Implementation 6: Evaluation (+ additional four units an “advanced course”)	1: Welcome 2: Humans: some basics 3: User-Centered Design 4: WIMP and GUI 5: Post-WIMP 6: Future
Types of assignments	Quizzes Design assignments Peer evaluation and feedback	Individual exercises and group discussions, posted in online discussion threads	Quizzes Structured assignments Design assignments
University credits	Varies by location[1]; Statement of Accomplishment provided.	No.	Yes.
Completion	Various completion levels Apprentice track, Studio track or Studio Practicum	Activity-based (answer quality not assessed) OpenBadge.	Activity in submitting assignments, quality of assignments; Exam

[1] For instance, the Stanford HCI course that started in April 2013 can be taken for credit at the University of Helsinki, see `http://www.cs.helsinki.fi/node/72025`

All these courses are open in such a way that the educators cannot assume any specific learner profile. Even CUX is open since it is offered to all international exchange students who arrive at our university and sign up to take it. No background studies are required. This way, it is open to all interested, but in a much smaller scale. The students need to have study rights at our university to enroll.

All courses depend on student activity during the course weeks. For CUX only students passing the final examination get university credits. The examination is a closed book exam arranged online, but in a controlled environment where the students' identity can be verified. This is a key issue for MOOCs where cheating is a concern [6].

In the Stanford course, students were expected to give feedback on each others' design assignments. Assessing the quality of creative work and user interface designs is a challenge to students and educators alike [4], and it requires carefully designed assessment rubrics [14] that need to be adapted to each assignment.

While online courses bring flexibility of time and place, time management needs special attention. In all three courses, time management has proven difficult. If the schedule is set too tight, there will be drop-outs. For many MOOC students, the HCI course is an extra-curricular course, and other obligations are more pressing at times. Time management is especially challenging in those courses where the students are expected to work together, for instance, to discuss something together in small online groups. Some students have moved on already while the others are just beginning their work on the topic, which might be a reason for the rather small number of discussions in the currently ongoing HCICourse [8]. The deadlines of assignments need to be announced at the outset of the course so that it is possible to plan ahead.

The course administrators would benefit from knowing who will finish the task at some point and who in time, and who will drop out without submitting any answer. This is one area of empirical research in the Stanford course [3]. Student profiling can support learner-centered design of the whole VLE [18].

4 Videos as Study Material

Next we look at how the learning units (compare with Table 1) are delivered.

All three HCI courses offer many short (typically 10-20 minute) recordings where the educator gives a lecture with slides and demo material. Only two of these courses include video recording of the educator in front of a camera. The quality of the recordings varies – seemingly simple issues such as where the video camera and speaker's monitor are located matter in the end result (Figure 1). In CUX, the whole course contains only one welcoming video of the educator, to give a face to the voice. All other lecture slides contain audio only. This was an intentional design decision at the outset of building the course for two reasons. Videos of talking heads do not seem to provide added value content-wise. Changing the audio by re-recording is much easier than editing video content.

The personality of the educator comes through in the video to the students, and body language enhances the message. The appearance and informal outfit give a more lively experience than was achieved in early recorded lectures, such as those by Day and Foley [7]. The slides can be showing on a background screen, as a background window where the video overlay of the presenter is inserted, or as the single video window. This makes the lecture more dynamic than it would be if showing just the face or upper body of the lecturer and a static arrangement of the scene.

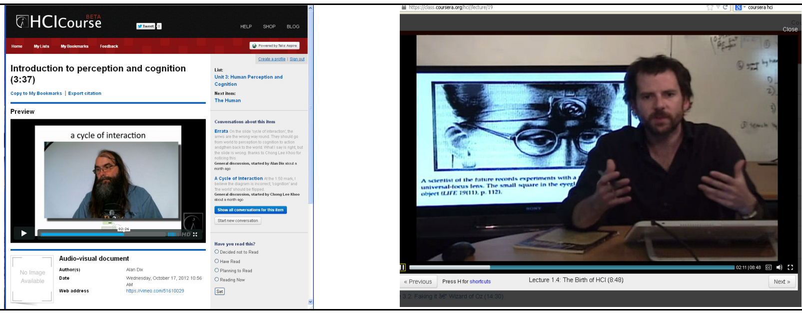


Fig. 1. Example screen shots of HCI course videos by Alan Dix (left) and Scott Klemmer (right), presented with their permission.

A live video stream, if not focusing on the lecturer's face, also enables showing demos and live sketching or scribbling. These are valuable assets in teaching about interaction. In Classroom 2000 [1], the lectures were captured with audio and scribbles made during the lecture became a part of the material. In this way, some of the classroom interaction was delivered to students who could not take part.

Fortunately, plenty of online video material exists in HCI in open access repositories[2] that can be directly linked to from the course platforms. This is done extensively in the HCICourse [8] and in CUX, too.

5 Study Methods: Making Notes, Working on Assignments

Compared to normal class long recordings of the Classroom 2000 [1], the MOOC videos are short, but still they seem separated from the exercises. A more directly interactive learning session by interleaving video and quizzes would enable faster learning cycles. In HCI education, "theory" and application interleave [7].

One of the first feedback messages in CUX was that a student asked for the PDF versions of the lecture presentation material used in the recordings to support note-taking while watching the presentations. Some students study by scribbling notes on the presentation [1]. The video recordings do not allow for this kind of capture of one's own content in the right context, and making notes on a fast-paced video lecture where only some parts of each slide are explained in depth would be difficult.

In the HCICourse [8], a student can make bookmarks, build lists of items, and make notes on various things. There is also a possibility to add discussion threads to

---

[2] A couple of the first online HCI video repositories:
`http://hccedl.cc.gatech.edu/`, `http://www.open-video.org/`,
`http://www.cs.umd.edu/hcil/pubs/video-reports.shtml`,
`http://www.acm.org/uist/archive/`

any type of resource or assignment of the course, such as the video in Figure 1. This way the student can ask for clarifications or comment on the presentation in context. However, spreading the questions and answers into several locations makes it more difficult to see the big picture and where the action is. This resembles the networking ideal of the first MOOC [5]: study materials and information do not need to reside in one web platform but can be distributed, and the wisdom of the crowd is utilized in assessing the value of the materials.

Collaboration between peers is often considered valuable in fostering learning [12]. CUX does not support peer interaction elsewhere but in the two discussion forums and the Comments block that is available in our Moodle 2.3 installation. To our surprise, the students did not use the forums for discussion, but the informal Comment block was often used for pointing out interesting links to peers.

Most assignments in CUX include quiz questions, such as in Figure 2. However, to enhance deep learning, in most occasions we asked for explanations and reasons for the answer the student gave. Often in HCI the answer is “It depends…”!

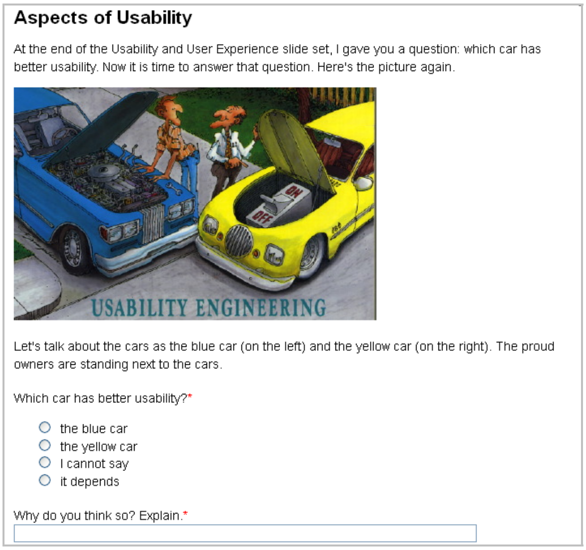


Fig. 2. A screen shot of a quiz page in CUX. The original car picture is from [17]

Before giving out the two design assignments of CUX we asked CUX participants about their desire to collaborate. The opinions were nearly unanimous: 81% of the respondents accept and are happy with individual study as the course study method. However, they want tutor feedback on their answers, so that they know if their answers are correct or not. Providing such feedback for design assignments and open questions (such as the question posed in Figure 2) is not trivial in massive courses, but appropriate for the small scale CUX. In the Stanford course that applies many extensive assignments, the educators have worked on assessment rubrics that enable student self-assessment of their assignment submissions [16]. The early results point out that the students improve in their assessment skills so that their assessments more accurately match assessments done by staff members. Working on the rubrics and assessments also improves teaching and makes the grading process transparent.

6 Discussion and Future Work

Instead of adopting a MOOC in full, a better pedagogy might result from combining the open access materials somehow to classroom tuition. One of the recurring themes with these open access video-based lectures is “flipped classrooms” [16]: the students are expected (and motivated [7]) to watch a recorded lecture video and work on relevant learning tasks already before class. This way the class time would enable more interaction with the students instead of the lecturing [4].

While the three HCI courses we compared differ in their methods quite a lot, they share similar challenges: keeping up motivation and interest, enabling time management, building on the principles of HCI and educating the students about our field.

The feedback from CUX participants has been encouraging. The online course format has been found good: the students enjoy being able to study when they have suitable time slots, control their own pace and plan ahead for the coming weeks. Nearly all who worked on the assignments also passed the examination – less than 10 percent were drop-outs. Many students said that they enjoyed sketching and learned most from the design tasks, but we noticed that the designs in some cases were not so good. It is evident that the feedback on their design ideas is essential for learning.

Our initial experiences also point out areas of improvement. Though the lecture presentations are shorter than in-class lectures on the same topics, they miss much of the dynamic interaction between the educator and the students. A shorter linkage between the “theory” of the lecture and the practical applications in the assignments would be needed to provide more reflection on the topics discussed in the lecture.

It is important to collect and share experiences of MOOCs in general and HCI MOOCs in particular. Teaching and study methods are evolving, and materials and solutions developed for MOOCs can help local efforts like ours. All in all, CUX has so far served its purpose well. Although the development effort was extensive, it provides a platform for individual study throughout the year, and separate course instances can be offered repeatedly at a low cost per instance.

Acknowledgements. Professors Dix and Klemmer are thanked for feedback on an early version of this paper. Ville Hämäläinen, CUX teaching assistant, was indispensable in running the course for the first time, and we thank the CUX course participants for filling in many questionnaires providing us with a glimpse of the experience.

References

1. Abowd, G.D., Atkeson, C.G., Feinstein, A., Hmelo, C., Kooper, R., Long, S., Sawhney, N., Tani, M.: Teaching and Learning as Multimedia Authoring: The Classroom 2000 Project. In: Proc. MULTIMEDIA 1996, pp. 187–198. ACM Press, New York (1996)

2. ACM Education Board/Council. Online Learning in Computing. ACM Whitepaper (2013), `http://www.acm.org/education/online_learning_white_paper.pdf`
3. Cheng, J., Kulkarni, C., Klemmer, S.: Tools for predicting drop-off in large online classes. In: Proc. Conf. Companion CSCW 2013, pp. 121–124. ACM Press, New York (2013)
4. Cooper, S., Sahami, M.: Reflections on Stanfords's MOOCs. Comm. ACM 56(2), 28–30 (2013)
5. Cormier, D.: What is a MOOC? (2010), Video available at `http://youtu.be/eW3gMGqcZQc`
6. Daniel, J.: Making Sense of MOOCs: Musings in a Maze of Myth, Paradox and Possibility. Journal of Interactive Media in Education 18 (2012), `http://www-jime.open.ac.uk/jime/article/viewArticle/2012-18/html`
7. Day, J., Foley, J.: Evaluating a Web Lecture Intervention in a Human-Computer Interaction Course. IEEE Transactions on Education 49(4), 420–431 (2006)
8. Dix, A.: HCI Course (2013), `http://hcicourse.com`
9. Dyson, M.C., Campello, S.B.: Evaluating Virtual Learning Environments: What Are We Measuring? Electronic Journal of e-Learning 1(1), 11–20 (2003)
10. Hardesty, L.: Lessons Learned from MITx's Prototype Course. MIT News Office (2012), `http://web.mit.edu/newsoffice/2012/mitx-edx-first-course-recap-0716.html` (July 16, 2012)
11. Hyman, P.: In the Year of Disruptive Education. Comm. ACM 55(8), 20–22 (2012)
12. Kester, L., Kirschner, P., Corbalan, G.: Designing Support to Facilitate Learning in Powerful Electronic Learning Environments. Computers in Human Behavior 23, 1047–1054 (2007)
13. Klemmer, S.: Stanford Course "Human-Computer Interaction" in Coursera (2012), `https://class.coursera.org/hci/`
14. Kulkarni, C., Klemmer, S.: Learning Design Wisdom by Augmenting Physical Studio Critique with Online Self-Assessment. Technical Report, Stanford University (2012)
15. Leber, J.: In the Developing World, MOOCs Start to Get Real. MIT Technology Review (March 15, 2013)
16. Martin, F.G.: Will Massive Open Online Courses Change How We Teach? Comm. ACM 55(8), 26–28 (2012)
17. Nielsen, J.: The Usability Engineering Lifecycle. IEEE Computer 25(3), 21–22 (1992)
18. Notess, M.: Usability, User Experience, and Learner Experience. eLearn Magazine (2001)
19. Petre, M., Carswell, L., Price, B., Thomas, P.: Innovations in Large-Scale Supported Distance Teaching: Transformation for the Internet, not just Translation. In: ASEE/IEEE Frontiers in Education Conference 1997, pp. 332–338. IEEE (1997)
20. Petre, M., Shaw, M.: What's the Value Proposition of Distance Education? ACM Inroads 3(3), 26–28 (2012)
21. Richards, D., Kelaiah, I.: Usability Attributes in Virtual Learning Environments. In: Proc. of the 8th Australasian Conf. on Interactive Entertainment IE 2012. ACM Press, New York (2012)
22. Russell, D.M., Klemmer, S., Fox, A., Latulipe, C., Duneier, M., Losh, E.: Will Massive Online Open Courses (MOOCs) Change Education? In: Extended Abstracts on Human Factors in Computing Systems (CHI EA 2013), pp. 2395–2398. ACM Press, New York (2013)
23. Sutcliffe, A.G., Alrayes, A.: Comparing User Experience and Performance in SecondLife and Blackboard. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) INTERACT 2011, Part III. LNCS, vol. 6948, pp. 680–696. Springer, Heidelberg (2011)

Audvert: Using Spatial Audio to Gain a Sense of Place

Liam Betsworth[1], Nitendra Rajput[2], Saurabh Srivastava[2], and Matt Jones[1]

[1] Future Interaction Technology Lab, Swansea University, SA2 8PP, UK
{csliamb,matt.jones}@swansea.ac.uk
[2] IBM Research, 4, Block C, Vasant Kunj, New Delhi – 110070, India
{rnitendra,sausriva}@in.ibm.com

Abstract. We introduce Audvert – a system that facilitates serendipitous discovery and navigation through spatial audio; used to navigate and discover points of interest in large, unfamiliar indoor environments. Our main aim was to create a lightweight spatial audio display that can convey a sense of a place without complex point and select interactions. We conducted a preliminary study comparing two audio types to see which best suited sound localization and a study of Audvert used in a real world scenario. Our findings suggest that long continuous audio performs better than short intermittent audio for sound localisation. We also discover a change in behaviour when using the system, with a large percentage of users wanting to visit newly discovered shops after using the system. We discuss the findings and draw research conclusions.

Keywords: Non-visual Navigation, Wayfinding, Auditory Display.

1 Introduction

When a visitor enters a shopping mall for the first time, it is likely that they do not know where they are going. To maximize the available space in a mall, shops are often packed closely together, and depending on the specific layout of the shopping mall and the location of the visitor, it is also likely that not all shops are visible. All of these factors can make a large, unfamiliar indoor location a confusing place to be.

Previous research on users' use of technology in unfamiliar places has shown that tourists often exhibit a wandering behaviour and then home in on serendipitous discoveries [1]. To aid such activities, maps (physical and digital) are sometimes used, although as mentioned by McGookin et. al [8], maps are not best suited to these kinds of activities due to changing contexts and the need to constantly refer to a map. We argue that tourist behaviour can be linked quite closely to visitor behaviour when in large and unfamiliar places such as shopping malls. It is easy to imagine a visitor to a shopping mall, wandering, homing in on anything of interest to them.

In an effort to aid such visitors, we have developed a system that attempts to facilitate navigation (homing in) and serendipitous discovery (wandering) using only spatial audio. Audvert is a mobile application that attempts to give a user a sense of place. Conventional maps are good at summarising what points of interest are available to a user in the nearby environment, but Audvert attempts to give the user a sense

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 455–462, 2013.

of what is actually around them by delivering spatial audio feedback. When a user hears a point of interest (POI) played back through their headphones, the audio sounds as though it originates from the real physical location of that POI. A visitor may not always be able to see a shop, but with Audvert they can infer the direction and proximity from the audio feedback, enabling them to then look around to see possible pathways of getting there. Where existing audio navigation and discovery systems have simply focused on alerting a user to the proximity or direction of a POI [4, 5, 7, 8], Audvert is unique in that it is used indoors, layering useful, spoken information about points of interest with directional and proximity feedback. Audvert is a lightweight system that offers the user continual feedback, allowing the user to engage when they want. The following scenario, along with Fig. 1, illustrates how a visitor might use Audvert in a shopping mall.

Scenario. *Fred is at a large shopping mall but is unsure about which shops to visit. Putting on his headphones and opening the Audvert application, he holds his phone out in front of him **(1)**. The Audvert app randomly picks a store from the mall and begins playing an audio clip containing the name and description of the goods and services available there. Fred can hear this information on his right side indicating that this particular store is in that direction **(2)**. Deciding he is not interested in this shop, Fred waits a few seconds for the next random selection to be played. After listening to the name and short description of the second store, which is now being played in his left ear, Fred decides he wants to know more and signals this via a shake gesture with his phone **(3)**. Fred now hears additional information and uses the direction the sound is coming from to guide him to the store **(4)**. As he approaches the store, the amplitude increases.*

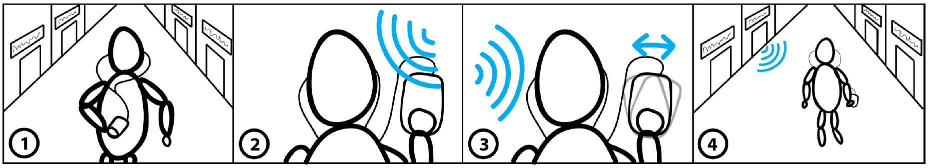


Fig. 1. User interaction in Audvert (explained in scenario above)

2 Related Work

As inexpensive, powerful smart phones have become more popular, soundscape and auditory display technologies have reached the consumer domain. A large amount of research on these topics has focused on their deployment for museum exhibits [3] and outdoor, augmented-reality soundscape experiments [2]. In terms of its applications in way finding, earlier work has tended to focus on the audio modality as a basic means of A to B navigation [4, 5], but more recently, researchers have focused on the specific problem of tourist way finding. McGookin et. al developed a system named Audio Bubbles [8], "employing non-speech audio to support tourist way finding." The researchers argue that the way in which tourists navigate is not so well supported by traditional visual maps. We argue that this is also the case for unfamiliar *indoor* environments such as shopping malls. Our system similarly attempts to facilitate serendipitous discovery – something that visual maps do not focus on.

Audio Bubbles uses a Geiger counter metaphor to signify proximity to a POI, similar to the early audio navigation system, AudioGPS [4]. In our design, we have chosen against using this common geiger counter metaphor. Instead, we use amplitude to give feedback of proximity. We believe this is a more natural mapping, as sounds naturally get quieter as they are further away. Unlike Audio Bubbles, we also incorporate directional feedback. We encode directional and proximity feedback in our audio recordings at runtime, similar to ONTRACK [5], although using the full 360° for direction instead of simple left and right panning. By not using the Geiger counter metaphor, we are also able to use audio recordings such as shop descriptions.

Different kinds of audio feedback for auditory displays have been explored in depth by Vazquez-Alverez et. al [13]. One auditory display system that uses spoken word is PULSE [7]. PULSE turned local Twitter tweets into spoken word using speech synthesis. When users approached the physical location where a tweet was composed, it would play to the user. Audvert is again, different from PULSE as we use spatial audio that takes distance into account. PULSE was designed to provide the user with a social vibe – "an intrinsic understanding of the people, places and activities around their current location." We strive to achieve a similar sense of place in Audvert, but through using pre-recorded audio content relevant to the shopping mall.

One problem that can occur when too many spatial audio streams are playing concurrently is increased cognitive load [12]. To combat this problem, we have chosen only to play one spatial audio stream at a time. In doing this, we also make the process of stream selection easier for the user. Most audio displays make use of a bearing-based selection technique (e.g. [6]), but this technique has its problems. When the number of selectable entities (audio sources in this case) increase, the bearing gap between each decreases. According to Strachan and Murray-Smith [11], effective selection was maintained down to around a 20° separation with a 10° target width. As Audvert only ever uses a single audio stream, we use a simpler interaction technique – a shake gesture to select or deselect the audio stream. We have chosen to use a gesture over a button, as previous research on gestures as a means of control for mobile devices has shown that they are a reliable, eyes-free interaction technique [9]. Further research has also been carried out on the practical applications and social acceptability of mobile gestures in the real world [10]. We believe the shake gesture is a simple and discreet gesture that most users will be able to use and will be comfortable with.

In summary, Audvert is novel in that it is an attempt at indoor audio navigation and discovery, not only providing proximity and directional feedback like previous systems, but also giving useful spoken information. It also tries to simplify the task of spatial display selection by making only one choice available at a time.

3 Audvert

Audvert is a prototype system designed to give users a better sense of the place that they are in using only spatial audio. To use the system, a pair of headphones and the

Audvert application is required. When using Audvert, sounds appear to originate from their actual physical locations. A user should be able to infer proximity through amplitude (closer is louder) and direction through the panning of sounds. Spatial audio feedback in the system is constantly changing due to lightweight interactions such as the user walking around (changing their location in the system) or the user changing direction (changing their listening orientation). Audvert uses a digital embedded compass, so as the user rotates, sounds continue to appear as though originating from the same physical location in space. With regards to location tracking, we decided upon a Wizard of Oz study methodology for Audvert. Users are not told that their location is entered manually elsewhere, so it is fair to say that users believe that their location is being automatically tracked. We use a separate purpose built system where a researcher may tap a map to update the location of a user. This Wizard of Oz approach was chosen as it offers a high accuracy with little complication. GPS was not a viable option as the system was used indoors. In reality, this kind of system would most likely work indoors with WiFi localisation.

When considering the user interaction with the prototype system, we decided that it should facilitate three separate functions, all relative to tourist way finding [1] - to aid serendipitous discovery of POI (wandering), to deliver useful information about them and to offer users the ability to navigate towards them if they wished to do so (homing). To facilitate serendipitous discovery, Audvert randomly chooses points of interest in the local environment, playing them back one at a time to the user. Useful information about the points of interest is spoken in the audio recordings (sales, offers and general information about the shop). Once a user has selected a POI, Audvert facilitates navigation through continually playing spatial audio for that POI.

4 User Evaluation

Two studies were conducted to test Audvert. The first was used to determine audio preference and performance. The second study gave participants the opportunity to use the system fully as they would do in a real life scenario. Both studies were conducted in a shopping mall in New Delhi, India.

Prior to the evaluation, early testing with users alerted us to the issue of users pointing the device in a different direction to their orientation. Because of this, a static arrow was added to the GUI as an affordance, so that users would remember to point the device in front of them at all times. There were also instances where users would forget the name of the shop they were currently listening to. For this reason, the name of the current shop the user is listening to was also added to the interface.

4.1 Preliminary Study: Cube Placement

Before embarking on a usability/user experience study of Audvert, we wanted to ensure we picked the optimal type of audio for use within the system. With this in mind, we recruited 24 participants (18M, 6F, 16-37 years) for a between-subjects study testing the effectiveness of two different types of audio; short intermittent and long continuous. We define short intermittent as audio clips that do not have continuous background sound and are short in length (i.e., purely spoken word, 5-6 seconds long). Long continuous clips contain continuous background sound and are longer in length (i.e., spoken word with a musical backing track, ~25 seconds long).

Our goal in this preliminary evaluation was to determine which of these two types of audio provided users with the most accurate path to the audio source location. This was achieved by playing five audio sources and asking each participant to place five paper cubes on a map of the shopping mall that best matched the location of where sounds were originating from. The map was A3 in size and each cube was $2cm^3$. Direction scores were calculated depending on the cube placement angle of error. Scores from 10 to 0 were given for 22.5° segments. Proximity scores where calculated depending on the cube placement error in centimeters (distance from user). None of the participants we recruited had any prior knowledge of the shopping mall.

The results of this evaluation showed that on average, participants using long continuous audio performed better in terms of perceiving both proximity ($p < 0.0068$; Unpaired t-test; 1.8cm average difference) and direction ($p < 0.0018$; Mann-Whitney; 11.4/50 average score difference). We therefore decided that short intermittent audio should be used as a brief alert of a shop's existence (when a shop is unselected), and long continuous audio should be used to give a person more information about a shop and allow a user to navigate towards it if they wish (when a shop is selected).

4.2 Usability and User Experience Study

The purpose of this study was to gain a better understanding of the usability and user experience offered by the system. We wanted to discover if Audvert could be used to navigate towards a shop that a user does not know the location of, and more generally, how a user feels about their surroundings when using the system. In search of answers, we attempted to evaluate the system in a real usage scenario, such as when a visitor is walking in a shopping mall and wants to know more about the shops there.

Participants. Participants of this study were again, real visitors to the shopping mall. We recruited 19 participants (10M, 9F, 17-47 years), none of which had taken part in the preliminary study. Approximately 3 were regular visitors to the mall. All participants could speak and understand English. No participant reported hearing problems. All participants had used a mobile phone before, but not all had used a smartphone.

Task and Measurements. Participants of this study were expected to explore the shopping mall using the system. Each participant was personally allowed to choose what he or she used the system for. Qualitative data on user experience with the system was gathered through a post-study interview. Questions asked in this interview were: Has the system given you a better or worse understanding of the shop locations? How would you describe your sense of place after using the system? *(Level of comfort and familiarity with the place)* Did you hear any shops that you have not visited before? *(Would you consider visiting one of them now?)*

Procedure. Participants were recruited individually as they walked through the shopping mall and were given a short introduction to the Audvert application. Participants were asked to explore the shopping mall using Audvert. As participants moved around the shopping mall, a researcher stood around 20 meters away and updated the participant location every few seconds. While exploring, participants were free to interact with the system as they wished. At the end of each study session, the participant was asked a series of questions in a post-study interview regarding their experience with the system. No incentive was given for completing the study.

Results. The following section details the responses given to the questions asked in the post-study interview. There were some participants that could not decide upon an answer to some of the questions. These have been included as N/A in the results. Percentages are relevant to the number of participants that answered each question.

	Response			
Question	**Better**	**Unaffected**	**Worse**	**N/A**
Has the system given you a better or worse understanding of the shop locations?	79% *(15)*	5% *(1)*	16% *(3)*	0
How would you describe your sense of place after using the system?	80% *(12)*	7% *(1)*	13% *(2)*	(3)
	Yes	**Maybe**	**No**	
Would you consider visiting a new shop you discovered using the system?	63% *(10)*	31% *(5)*	6% *(1)*	(3)

Table 1. Quantitative results from questions asked in the study

Informal Observations & Feedback. It became apparent in the study that not every participant was selecting a shop by shaking the device. By avoiding this function, the participant was not able to gain extra information about a shop (a long description) – the system however, is still usable without this function. As Audvert provides no visual feedback of where the user is in the system, one participant was unsure whether a shop was selected or unselected at any point in time. This participant carried on for a further two minutes and after this period appeared to be much more comfortable with the system. One participant that successfully used the shake gesture to activate a shop was seen attempting to navigate towards the selected shop. This participant attempted to follow the direction of sounds, and did so successfully by finding the shop. The participant mentioned that they knew they were going in the correct direction because they could hear the sounds getting louder as they got closer to the shop. This person said that they had never visited the shop before, and that they had no prior knowledge of its location. There was also another situation where a participant used the varying volume of sounds to navigate towards a shop. This participant had selected a shop, walked in one direction and then realised that the volume was decreasing - inferring that they were heading in the wrong direction. The participant then turned around to face the correct direction and successfully found the shop that they had selected.

4.3 Discussion

Most participants enjoyed the experience, claiming that the system appeared to be ‘very accurate’ and that the direction sounds came from was the same direction that they could see the shops. A large portion (79% (15)) of participants said that they had a better understanding of the location of shops after using the system. One participant who claimed a worse knowledge of shop locations said that it was “difficult to differentiate between the distances of each shop.” Another participant agreed, finding it “difficult to tell the distance to a shop”, but claiming the “direction was good.” Although 79% of participants said that their understanding of shop locations had improved, 1/3 of these (5 participants) claimed that they could either infer direction or distance, but not both. In most of these 5 cases, participants could only infer direction (4 participants). This is an interesting result, as without knowing both, one would imagine it being difficult to pinpoint an exact location. These participants obviously believed that knowing only direction or proximity was sufficient for knowing the location of a shop. This strengthens the argument for spatial audio as a navigation aid.

A similarly high percentage (80% (12)) claimed that they had a better sense of place after using Audvert. There were 5 participants who did not answer this question, unable to differentiate between knowing the locations of shops and a sense of place. One participant said that the system gives a “good sense of the area. I could tell what kinds of shops and products were available.” Another participant argued that the system gave a bad sense of place as there were “lots of shops and I only heard a few.” Although a similar percentage said they had a better sense of place and an improved understanding of shop locations, not every participant who had an improved understanding of shop locations had a better sense of place, showing that the two weren’t regarded as the same question by participants.

80% of participants heard a shop they had not visited before, of which, 63% (10) said they would definitely consider visiting that shop and 31% (5) said they might visit now. There were 4 participants who had previously visited every shop that they heard when using the system. Only one participant decided that they would not visit a new shop they had discovered, saying that they are the “kind of person that comes to the mall for one thing.” There does appear then, to be a compelling change in behaviour, where the system seems to encourage participants to visit new, serendipitous discoveries that are made. Generally, most participants explained that they would like to visit these shops because of sales and discounts mentioned in the audio recordings. Participants who said they might visit a new shop often said that they would do so if more detailed information about a shop was available. Judging by this feedback, it seems that the content is an integral factor of this change in behaviour.

The majority of participants liked the shake gesture. Two participants explicitly mentioned that they would “prefer the shake over a button.” In contrast, two participants said that they were not comfortable with the shake gesture, preferring a button. One participant’s concern came from not wanting to shake their expensive mobile phone. To satisfy all users, a multimodal approach could be introduced, where the user chooses whether the interaction occurs onscreen or through a gesture.

5 Conclusions

We have introduced Audvert to illustrate the potential of a lightweight system that facilitates navigation and serendipitous discovery of large indoor spaces using spatial audio. While the idea of an audio display for navigation and discovery is not a new one, we argue that the novelty of our approach lies within its indoor application and the feedback and selection techniques employed. We use amplitude and directional feedback to manipulate useful, spoken audio information about a point of interest in real time. Audvert has also introduced a simplified selection technique for spatial displays through using a single active element and a simple shake gesture to select.

Our user evaluations have shown that Audvert can convey useful spatial audio feedback, and that it is capable of helping users navigate towards and discover new points of interest. We saw multiple examples of users successfully arriving at a shop and for the majority of participants, we witnessed a change in behaviour with participants claiming that hearing new shops made them want to go and visit there.

References

1. Brown, B., Chalmers, M.: Tourism and mobile technology. In: Eighth European Conference on CSCW 2003, Helsinki, Finland, vol. 1, pp. 335–354. Kluwer Academic, Dordrecht (2003)
2. Cater, K., Fleuriot, C., Hull, R., Reid, J.: Location aware interactive applications. In: ACM SIGGRAPH 2005 Sketches, SIGGRAPH 2005 (2005)
3. Heller, F., Knott, T., Weiss, M., Borchers, J.: Multi-user interaction in virtual audio spaces. In: CHI EA 2009, pp. 4489–4494 (2009)
4. Holland, S., Morse, D.R., Gedenryd, H.: Audiogps: Spatial audio navigation with a minimal attention interface. Personal Ubiquitous Comput. 6(4), 253–259 (2002)
5. Jones, M., Jones, S., Bradley, G., Warren, N., Bainbridge, D., Holmes, G.: Ontrack: Dynamically adapting music playback to support navigation. Personal Ubiquitous Comput. 12(7), 513–525 (2008)
6. Marentakis, G.N., Brewster, S.A.: Effects of feedback, mobility and index of difficulty on deictic spatial audio target acquisition in the horizontal plane. In: Proceedings of CHI 2006, pp. 359–368 (2006)
7. McGookin, D., Brewster, S.: Pulse: the design and evaluation of an auditory display to provide a social vibe. In: Proceedings of CHI 2012, pp. 1263–1272 (2012)
8. McGookin, D., Brewster, S., Priego, P.: Audio bubbles: Employing non-speech audio to support tourist wayfinding. In: Altinsoy, M.E., Jekosch, U., Brewster, S. (eds.) HAID 2009. LNCS, vol. 5763, pp. 41–50. Springer, Heidelberg (2009)
9. Pirhonen, A., Brewster, S., Holguin, C.: Gestural and audio metaphors as a means of control for mobile devices. In: Proceedings of CHI 2002, pp. 291–298 (2002)
10. Rico, J., Brewster, S.: Usable gestures for mobile interfaces: evaluating social acceptability. In: Proceedings of CHI 2010, pp. 887–896 (2010)
11. Strachan, S., Murray-Smith, R.: Bearing-based selection in mobile spatial interaction. Personal Ubiquitous Comput. 13(4), 265–280 (2009)
12. Vazquez-Alvarez, Y., Brewster, S.A.: Eyes-free multitasking: the effect of cognitive load on mobile spatial audio interfaces. In: Proceedings of CHI 2011, pp. 2173–2176 (2011)
13. Vazquez-Alvarez, Y., Oakley, I., Brewster, S.A.: Auditory display design for exploration in mobile audio-augmented reality. Personal Ubiquitous Comput. 16(8), 987–999 (2012)

Rhythms of the Domestic Soundscape: Ethnomethodological Soundwalks for Phatic Technology Design

Hanif Baharin[1], Sean Rintel[2], and Stephen Viller[2]

[1] Universiti Teknologi MARA Malaysia, 40450 Shah Alam, Selangor, Malaysia
[2] The University of Queensland, Brisbane, QLD, 4072 Australia
ahmadhanif@perlis.uitm.edu.my,
s.rintel@uq.edu.au, viller@acm.org

Abstract. The importance of the domestic soundscape as a context for technological interventions has received little attention in HCI research. In this paper, we discuss how an ethnomethodological soundwalk method facilitated design principles for a phatic technology probe for seniors living alone. Taking soundscape concepts as a starting point, we suggest that the soundwalk works much like a breaching experiment, changing the participant's role in engaging with their soundscape from *reactive automatic agent* to *proactive reflective agent*. This enables participants to reveal their own systematic orderliness when accounting for everyday sounds. We find that sounds are accounted for in terms of people placed in narratives. As such, we argue that phatic technologies use new sounds and rhythms to augment the domestic soundscape to take advantage of people's abilities to create social narratives from limited cues.

Keywords: Domestic soundscape, soundscape study, ethnomethodology, breaching experiment.

1 Introduction: Phatic Audio Technologies

The importance of the domestic soundscape as a context for technological interventions has received little attention in human-computer interaction (HCI) research. In Australia, about 29 percent of people who are 65 years of age and over and 39 percent of people aged 85 years and over lived alone in 2006 [1]. Inspired by Tacchi's work [2] showing that the radio is used in domestic settings to negate the feeling of social isolation, we are exploring new design ideas for phatic audio technologies: lean audio signalling "awareness systems [that] contribute to a feeling of human connectedness" [3:178] through interventions into the domestic "soundscape" [4] that encourage social narratives. In this paper, we discuss an ethnomethodological [4, 5] version of Schafer's [6] soundwalk method in which we found that even when living alone, the domestic soundscape is adorned by the sounds of other people which are accounted for in terms of social narratives. The design implication of this finding is that lean signalling can be leveraged to encourage such social narratives for a range of use cases.

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 463–470, 2013.

1.1 Soundscape Approach in HCI

As opposed to visual displays, HCI design research on soundscape augmentation has been quite limited. This research, however, consistently finds that sounds are treated as accountable and narratable. Alexanderson and Tollmar [7:255] have explored a participatory design-oriented phenomenological approach to augmenting factory soundscapes, and found that "an important part of the design process is the operator's contribution to the concepts suggested." Oleksik, Frohlich, Brown and Sellen [8], adopted the soundscape approach to inform design in the domestic context. After conducting a "sound tour", in which participants showed researchers around their homes and discussed the sounds they heard, they found that sounds in the home had special meanings and functions for the inhabitants.

In The Tuning of the World, Schafer [6] provides a novel view of the problem of noise in society. Instead of providing ways in which noise may be abated, he argues that society as a whole should be actively designing sounds in the world to make it a sonically beautiful place. Schafer [6] argues that technology has created problematic changes to the ways humans appreciate sounds in their environment, coining the term "schizophonia" to propose the tendency of people to become distracted listeners due to the technological separation of sounds from sources. Although our research adopts Schafer's approach to encourage listening, we disagree with Schafer's entirely negative evaluation of technology. We take the stance that technology can help to create a more acoustically pleasing world in which listeners are encouraged to add context to dislocated sounds.

2 Ethnomethodologically-Informed Domestic Soundscape Study

We see our work as a continuation of that of Oleksik et al. [8], but with an ethnomethodological stance shaping our approach and interpretation of data. We were trying to elicit how people account for their experience of the domestic soundscape and how people achieved the common sense understanding of their soundscape. Garfinkel [4] argues that social order is not governed by theories unknown to social actors, but achieved by moment-by-moment effort taken by people to organise their behaviour. Ethnomethodological investigations aim to illustrate how people make sense of the rationalisation of actions by, and as displayed by, others — in other words "What everyone knows that everyone knows" [4:75].

Schafer [6] introduces the "soundwalk", in which participants take the researcher around their house and described the sounds they hear, as a listening exercise to increase people's awareness of their soundscape. For us, the soundwalk is a way to gather the foundational data about how people make sense of the types and rhythms of sounds in their domestic environments. Rather than simply moving from room to room and passively listening, in a soundwalk the researcher guides the participants' attention to the sounds in their environment by encouraging discussion of routines undertaken in each area of the environment. Eight participants (three males and five females) with ages ranging from 51 to 87 were involved in this study. The names of the participants have been changed to protect their privacy. All the participants were living alone in their own homes in suburban areas in Brisbane, Australia.

Our soundwalks included both researcher-guided and self-guided versions. In the researcher-guided versions, participants were given a sound recorder and a diary (for another parallel study) and were asked to record sounds they heard in and around their dwellings. The soundwalk was conducted either when the researcher first met the participants in their home or when the researcher visited them a week later to collect the diary and the sound recorder. Data from the soundwalk and diary were analysed to inform interview guides. During interviews, they were encouraged to reflect on the recording process and the sounds they recorded. For privacy reasons some participants conducted self-guided soundwalks. They were given an information pack containing a diary, a sound recorder, and stationery. They were instructed to draw a plan of their home, and go through every room and area, listing the sounds they heard based on their daily routine, and to take the sound recorder along with them to record the sounds they heard.

3 Soundscape Narratives and Rhythms

3.1 The Mundane Soundscape

For the participants there was nothing profound about the everyday sounds they heard, from within and outside the home, because the sounds were a part of 'normal life' (e.g. Janet). However, many of these participants expressed mundanity through evaluative paradigms: being alive or not (e.g. Janet) or whether sounds might be a cause for 'worry' or not (e.g. Ian). The evaluative nature of even baseline 'nothing special' responses is, we argue, a fundamental feature of the accountability of the soundscape.

> *Janet: Well you do know that life is normal. You're alive. Perhaps you could tell but you could be walking around like zombie not alive but if you don't hear the sound and the noises of everyday activity. You know you're conscious [laugh].*

> *Ian: So, I do hear noises, but they don't worry me. Some of them might mean something. Some of them don't.*

3.2 Sounds Mean People in Narratives

Our analysis of the participants' diaries and of the first few interviews revealed that all participants described other people as the producers of sounds in their soundscape. Despite living alone, other people's sounds were treated as decorating their soundscape.

Lucy: Eh but there is sound going all the time, I mean there are people coming and going, and that cat, I mean the door is not always open but you can...

Marge: Well, you feel like little people about, you know? Because he goes to work all day, this one's not there. And you don't sort of see these people in the complex very much, so you've got the sound of the children, which is quite nice.

These sounds about other people in domestic soundscapes are also talked about in the form of narratives. For example, Janet wrote in her diary:

Janet (Diary): Noisy street, people coming back from stadium to collect their cars.

When Janet wrote this, she was not describing what she heard as discrete sound stimuli in their objective forms, such as footsteps on asphalt or engines being ignited. Rather, she formulated the sounds in terms of other people undertaking social activities with narrative elements — places, characters, and events — and narrative structures — time and reasoning. We found many instances where the participants talked about their domestic soundscape in terms of other people's actions. As shown in bold in the transcripts below, a similar structural pattern of expression was found in the responses of many participants: The sound event was first presented in its narrativised form, and then reformulated in terms of the discrete objective sound stimulus:

Roseanne: Right, the only other sound I hear from this area, and when the windows are open I might ***hear my neighbours showering. Or I hear the shower.***

Karl: Yes actually the people at the back have ***some teenage boys and sometimes they'd kick a football around*** *and you hear* ***the sound of football being kicked*** *and you know that's my neighbour at the back.* ***The kids are playing.***

Lucy: You know, I mean it – you can't help it here, I mean you can hear ***when other people go to the toilet*** *because the toilet, well sometimes you can hear when* ***the toilet flushes****.*

It is not within the scope of this paper to speculate why this pattern was so commonly employed. The more important point is to notice that the process of undertaking the soundwalk provided a way for participants to endogenously notice their own methodical practice of accountability, a reflexive version of what Garfinkel (after Mannheim) referred to as the "documentary method of interpretation" [4:78] .

3.3 Soundscape Rhythm

Similar to Oleksik et al. [8], we found participants contextualising sounds as chronological aggregates: rhythms. These rhythms were used in narratives, usually coalescing around the meaningful activities associated with a time of day. Rhythms in the domestic soundscape were especially focused upon the repetition of time-heralding sounds, such as radio programs, clock chimes, and church bells:

> *Lucy: Well I know what programs are on, I mean I listen to Radio National most of the time, and I would know usually what programs are on what time. I mean I mightn't know precisely to the minute because they rarely announce the time. But I know if it were between nine and ten, because ah, between nine and ten there's "Life Matters", after ten, there's a, whatever they call it "Book review" or something […].*

> *Ian: Yes, way up in the back a long way up the hill a church bell rings at five o'clock every night.*

Participants also learned rhythms and developed narratives about nearby houses. Ian knew how often the house near his was occupied based on the sounds made by its occupants and Karl recognised that during the renovation of his neighbour's house the noise started at seven every morning:

> *Ian: No, it's just background noise, it's there, you know something's going on. There's nobody there very much during the day. Even less on that side**, people come down about once a month. Stay there three days. And I can hear when they're home**. Because they've got wooden floors [laughter].*

> *Karl: The exhaust fan [of my neighbour] I would normally hear in the kitchen or in the launch room because it's about there, and the renovations mainly I would hear them here, because they start in the morning, **they start about seven.***

The ability to learn and draw meaning from soundscape rhythm is most clearly shown when the participants discuss variation or disruption of the learned rhythm. Janet noticed her clock was not working properly, when it did not produce the same rhythm:

> *Janet: You know what I mean. It gives one on the quarter, two on the half and then you got the fourth on the, you know so. As I say, I couldn't tell you whether it chimed or not because it's a noise that must happen. I only know when something's gone wrong you know.*

Roseanne could discern the changes in her neighbour's routine based on the change of her soundscape rhythm. She knew that there were more sounds coming from her neighbour's house during weekends than during weekdays. When she did not hear them that weekend, she speculated that they were away.

> *Roseanne: Because the young people who live there are at work and look I haven't heard a sound from them the whole weekend. They may have been away. Just occasionally once **every three months or something they might have friends over in three o'clock in the morning**. I wish they'd go home. But it doesn't happen very often so you know you think well okay.*

4 The Ethnomethodological Soundwalk: The Soundwalk as a Breaching Experiment

This section highlights our contribution to HCI by suggesting that the soundwalk may be treated as a form of breaching experiment. The participants experienced the domestic soundscape as mundane immersion, using their domestic soundscape as a resource for "background expectancies as a scheme of interpretation" [4:36]. That being said, they could not readily talk about the ever-present everyday sounds in their domestic soundscape without at least some prompting. This is in accordance with the observation by Garfinkel [4] that people are able to take actions based on their "background expectancies", but are "at a loss to tell us specifically of what the expectancies consist. When we ask him about them he has little or nothing to say" [4:37].

Breaching experiments were originally introduced by Garfinkel [4:38] as "aids to a sluggish imagination" to make transparent the shared common sense understandings or accounts of everyday actions. For example, by changing the accepted role of self, from being a family member to a boarder, the shared understanding of the actions that make the role of being a family member accountable can be made clear. However, when introducing a breaching experiment in HCI, Crabtree [5] is of the opinion that the disruption element of Garfinkel's experiments is taken too literally, and is hardly necessary as a required element. Indeed, not all of Garfinkel's breaching experiments reported "bewilderment". Many showed attempts to incorporate unusual behaviour into an acceptable shared framework of understanding.

We argue that this is what happened when we asked our participants to do the soundwalk. These behaviours were unusual to them as far as their experiences in engaging with their domestic soundscape are concerned. Being involved in our study was much like being involved in a breaching experiment because the participants changed their role from being reactive automatic agents to proactive reflective agents.

Our immersion in the soundscape, and our ability to monitor everyday sounds based on selective hearing, make our responses to sounds seem automatic. Many participants talked about focusing on 'something else' other than the sounds produced by their domestic activity, for example Karl said:

> *Karl: I guess your main focus in on something else, like you might be focusing, I'm gonna cook my dinner now, I'm gonna do the dishes [...] Not thinking about the actual sound of the plates going in the water [...] you see, you don't notice that.*

However, the recording exercise changed the participants from reactive agents to proactive agents, actively seeking sounds to record, which in turn made them more reflective about their soundscape. This, in combination with soundwalk and interview, made the participants assume the role of proactive reflective agents and enabled us to decipher their common-sense understanding of their domestic soundscapes. Karl, Lisa and Roseanne elucidated this:

> *Karl: So it was good exercise [the recording exercise] I quite enjoyed it. Cause the more I thought about it, I thought, "Ah...what am I gonna do?" as each day went on I thought of more and more things to. But I*

listen to something "Ahh...that's a good one." and I might come back to it. Cause some things, some of the recording I might have been the third or fourth go.

Lesley: Every day there are sounds no matter where you go.[...] I mean I realized that I can hear the noise around now until I just stopped and listened to it. But I mean you along every day activities and probably a lot of sound you completely don't realize there have been made.

Roseanne: And if you hadn't asked me to listen, I probably wouldn't have even heard the hum of the traffic, probably because; maybe to my thoughts would be focus on something else.

Therefore, we suggest that the change of roles, as encouraged in the soundwalk may be used as a breaching experiment to elicit the understanding of sounds in HCI research. Researchers may strive to design activities for the participants, or methods that change the participant's role of being a reactive automatic agent in the soundscape to some other roles, for example, as in our case, to a proactive reflective agent.

5 Implications for Design

Our findings show that, although the participants were living alone, their soundscape was intrinsically bound and shaped by the presence of other people in their environment. We are not claiming that this is universal, but this is at least true for all of our participants who live in urban areas, close to other people's dwellings. There are sounds in the domestic setting that denote the presence of others, and people relate these sounds in the form of narrative, and sometimes create speculative narrative based on the limited information they get from the sounds. The sounds of other people's routines add to the rhythm of the domestic soundscape, and people have the skills to discern this rhythm and notice the variations in the rhythm.

As noted in the Introduction, this design implication of the soundwalk method, and our findings in this instance, is that limited stimuli are contextualised as narratives of self and others endogenously even if not always explicitly. Further, when their attention is drawn to their soundscape, people are able to reflexively apply a documentary method of interpretation to their own developing understandings. Therefore, our contribution lies in the suggestion that soundscape studies may be designed as a form of breaching experiment. Both findings, we argue, are fundamental to the implication that lean audio signalling can be leveraged to encourage social narratives. These social narratives could be used to in a wide range of technology interventions.

We suggest that new sounds and new rhythms may be designed to augment the domestic soundscape to take advantage of these skills. Since domestic soundscape rhythm is linked to the presence and routines of people in the environment, social narrative creation via rhythmic sounds may be encouraged by design. That being said, it also opens up design questions regarding sonic privacy and etiquette in a shared soundscape created by technology. In our own research, we have used these ideas to create a phatic telecare technology probe called SonicAir, which uses lean audio signalling to represent the actions of a remote friend or family member. When SonicAir connected the homes of two people living alone, they were able to develop an ambient sense of social connection from which both took comfort. We believe that representing a remote person using audio phatic technology could be applied alongside telecare monitoring technologies to provide some of the benefits of social relationships while allowing for continued independence.

Acknowledgements. We are grateful to the participants for their time throughout our soundscape study and associated research.

References

1. Australian Institute of Health and Welfare: Older Australia at a glance (fourth edition). Cat. no. AGE 52, AIHW, Canberra, Australia (2007), `http://www.aihw.gov.au/publication-detail/?id=6442468045` (viewed May 17, 2013)
2. Tacchi, J.: Radio Texture: Between Self and Others. In: Askew, K., Wilk, R.R. (eds.) The Anthropology of Media, pp. 241–257. Blackwell Publishers Inc., Oxford (2002)
3. Vetere, F., Smith, J., Gibbs, M.: Phatic Interactions: Being Aware and Feeling Connected. In: Markopoulos, P., De Ruyter, B., Mackay, W. (eds.) Awareness Systems: Advances in Theory, Methodology and Design, pp. 173–186. Springer, London (2009)
4. Garfinkel, H.: Studies in ethnomethodology. Prentice-Hall, Englewood Cliffs (1967)
5. Crabtree, A.: Design in the absence of practice: breaching experiments. In: Proceedings of the 5th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques (DIS 2004), pp. 59–68. ACM, New York (2004)
6. Schafer, R.M.: The Tuning of The World. Knopf, New York (1977)
7. Alexanderson, P., Tollmar, K.: Being and mixing: designing interactive soundscapes. In: Proceedings of the 4th Nordic Conference on Human-computer Interaction: Changing Roles (NordiCHI 2006), pp. 252–261. ACM, New York (2006)
8. Oleksik, G., Frohlich, D., Brown, L.M., Sellen, A.: Sonic interventions: understanding and extending the domestic soundscape. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2008), pp. 1419–1428. ACM, New York (2008)

“Roger that!” — The Value of Adding Social Feedback in Audio-Mediated Communications

Rahul Rajan[1], Joey Hsiao[2], Deven Lahoti[3], and Ted Selker[1]

[1] Carnegie Mellon University, Pittsburgh, PA, U.S.A.
[2] National Taiwan University, Taipei City, Taiwan
[3] Independent
{rahulraj,tselker}@andrew.cmu.edu,
{hideysmk,deven.lahoti}@gmail.com

Abstract. Losing track of who is in a conversation, and what is being said, is always a problem especially on audio-only conference calls. This paper investigates how domain-independent social feedback can support such interactions, and improve communication, through the use of audio cues. In particular, we show how an agent can improve people’s ability to accurately identify and distinguish between speakers, reassure users about the presence of other collaborators on the line, and announce events like entry & exit with minimum impact on users cognitive ability.

Keywords: Audio-mediated, Conference calls, Considerate, Social feedback.

1 Introduction

Communication is a type of social action [1]. It can be verbal and non-verbal in nature. From a suggestive glance to an admonishing tone, people rely on all sorts of cues to assess the situation and regulate their behavior. Particularly while collaborating, people orient themselves and coordinate in creating a shared reality. They engage in this process to seek understanding, and to be understood. Feedback is pivotal to this process, and it propels and directs further communications. It helps in creating a shared awareness and mutual understanding.

When the communication is mediated by technology there is a reduction in these social cues or feedback. This creates a sense of disengagement and psychological distance. It is interesting to note that both video and audio-only conferencing suffer from the attenuation of these cues, albeit in different measures [2]. Hence, we find that the popularity of their use lies on a spectrum depending on the situation, the participants, the nature of the task, and the social set- ting. For instance, audio conference calls are widely used in business meetings [3], whereas desktop videoconferences are more popular in personal settings [4]. Even so, the reasons for users choice and preference are nuanced and complex, involving multiple tradeoffs related to intrusion, amplification of inattention, and mobility.

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 471–488, 2013.

While there has been significant work and progress to preserve social cues in video communications [5], audio-mediated technologies have not received the same share of attention, and depend largely on visual cues like participant lists to buttress communications [6].

In this work, we explore how social cues can be restored on the audio channel, while addressing some of its most often cited drawbacks [7, 2, 6]. These are predominantly social in nature, and focus on the process of interaction. These include the ability to accurately identify speakers, the notion of personal space for remote participants, and the issues of awareness about the presence of other collaborators. We design different types of audio cues, and experiment with feedback and feedforward techniques to better understand how these might support human communication. Our goal is to build a considerate agent that would know when and how to apply these techniques appropriately. Audio interfaces work well in spite of arguably have the lowest bandwidth for natural synchronous communications between multiple people. The difficulty in improving these communications with an agent further loading this narrow channel should then be maximally hard, making it an ideal place to demonstrate the utility of an agent being considerate and appropriate [8].

A second reason to experiment with augmenting the audio space is that many discussions and meetings involve documents and physical artifacts that occupy users attention. This makes display space expensive, and switching between display views task intensive. Besides, the visual channel is not the best medium to convey awareness information because human visual field is limited to the frontal hemisphere, and the foveal region in particular. This creates inherent limitations in the use of visual displays, wherein the user must see and attend to the display. Noticing visual changes also gets harder as tasks get more demanding, or if the display is cluttered.

Thirdly, people can perceive multiple audio channels simultaneously, and do so with considerable ease, especially while listening to music. In particular, we have the perceptual ability to hone in on a particular channel while filtering out the rest, commonly referred to as the “cocktail party effect” [9]. Thus, the audio channel can be used as an effective mode of transmitting background information (e.g., [10, 11]). In addition, audio can be used for conveying temporal information like whether an activity is occurring right now, and when actions start and stop. It can also be used to indicate spatial and structural information, like where the actions are happening, the type of activity and the qualities of the action (e.g., [12–14]).

To evaluate feedback techniques for improving communications, we choose the multiparty audio conference call with a shared-screen as our setting. Previ- ous work demonstrates that considerate agent cues can reduce distraction, and help conference call participants equalize contributions [15]. This paper moves further to show that considerate agents like CAMEO can more generally make call participants aware of others on the line and allow them to focus better on the conversation. In particular, we focus on speaker identification, audio presence, and entry & exit announcements. To evaluate speaker identification methods, we had participants listen to a pre-recorded five-person conference call and answer identity-related questions. We found that while speaker identification is hard, audio augmentations or spatially arranging the different speakers, could aid with identification success. We then show that by feed-forwarding simple audio cues, users were more assured, and less distracted about the presence of other collaborators. Finally, we had participants engage in a memory game while subjecting them to three kinds of entry/exit prompts. We show that by making a more natural, and less syntactic utterance, participants made fewer errors on average.

2 Related Work

There is a rich body of work on the use of audio for user interfaces, which provides the foundation for our work of supporting social cues in conversation. We briefly review how audio interface design has evolved, and the sounds and techniques others have used to provide audio feedback and guide user interactions. We then cover how audio has been used in distributed settings to allow peo- ple to coordinate better, and to increase shared awareness of remote events and activities.

2.1 Auditory Interface Design

Audio interfaces largely use two types of non-speech cues, namely, earcons and auditory icons. Earcons are synthetic tones whose timbre, pitch, and intensity are manipulated, to build up a family of sounds whose attributes reflect the structure of a hierarchy of information. Since earcons are abstract, they require training and need to be learned to be effective. Auditory icons are a more focussed class of audio cues, which are carefully designed to support a semantic link with the object they represent, making them easier to associate. Furthermore, sounds can be perceptually mapped to the events they indicate using symbolic, metaphorical and iconic methods.

Soundtrack [16] was one of the first auditory interfaces to use earcons and synthetic speech. More recently, Rigas et al. [12] demonstrated the use of earcons to communicate information about the layout of a building. Four different timbres (piano, organ, horn, and clarinet) were used to communicate the sections of the building. Floors were communicated by musical notes rising in pitch. A single note was rhythmically repeated to indicate room number, and combination of timbres was used to indicate hallways. Users successfully located the rooms but were not able to interpret the different hallways, suggesting that combina- tion of two timbres created confusion. Early in our work, we experienced how an overloaded audio dimension could easily be created by assigning multiple tracks of an orchestra to each participant. We focus on methods that prevent such overloading in a single audio dimension.

SonicFinder [17] was the first interface to incorporate the use of auditory icons. A variety of actions made sounds in the SonicFinder, including the ma- nipulation of files, folders, and windows. SonicFinder also made use of dynamic parameterized sounds to indicate temporal and structural activity, like file transfers producing a continuous filling up sounds, and different files producing different pitched sounds. In our work, we seek to explore when we can use the intuitive semantic mappings of auditory icons, over the arbitrary symbolic mapping of earcons.

A number of other works show how audio interfaces can improve interactions. gpsTunes [18] focussed on using adaptive audio feedback to guide a user to their desired location. As the user gets closer to the target, the music gets louder followed by a pulsing track to indicate their arrival. Schlienger et al. [19] evaluated the effects of animation and auditory icons on awareness. They showed that the auditory icons were commonly used to notify a change, and to focus attention on the right object just before it changed. AudioFeeds [20] explored how audio can be used to monitor social network activity, and PULSE [21] evaluated how audio cues can be used to communicate the local social vibes as a user walks around. This paper shows that such indicators might work well and not interfere with a conference call.

2.2 Activity Coordination in a Distributed Setting

SoundShark [22] was an auditory interface extension of SharedARK, a multiprocess system that allowed people to manipulate objects and collaborate virtually. It used auditory icons to indicate user interactions and ongoing processes, to help with navigation, and to provide information about other users. Users could hear each other even if they couldn't see each other, and this seemed to aid in coordination. This work motivated the development of ARKola, a simulation of a soft-drink bottling factory [23]. Temporally complex sounds occupied differ- ent parts of the audible frequency spectrum, and the sounds were designed to be semantically related to the events they represented. Also, instead of play- ing sounds continuously, a repetitive stream of sounds were used to allow other sounds to be heard between repetitions. Gaver et al. observed that the sounds allowed the people to keep track of many ongoing processes, and facilitated collaboration between partners. Users were able to concentrate on their own tasks while coordinating with their partners about theirs, when sound was providing the background information. We seek to employ similar techniques to show how we might improve the process of audio communication itself.

The CSCW community has also paid attention to the use of audio in distributed workspaces. In a shared drawing environment, Ramloll and Mariani [24] played different sounds for different participants, and spatialized the sound in the 2D environment to help with location awareness. Participants complained that the spatial audio was distracting, but it provided them with information about others intentions which helped them with turn-taking. McGookin and Brewster [25] looked into audio and haptic locating tools as well, while extending their single user GraphBuilder to a multiuser interface. They found that shared audio helped in mediating communication, and served as shared reference points, allowing users to refer to events they couldn't see. Our work seeks to extend this to situations where the fact of a persons presence is crucial to the outcome.

2.3 Shared Awareness in a Distributed Setting

The Environmental Audio Reminders (EAR) system [10] transmits short auditory cues to people's office to inform them of a variety of events around their building . For example, the sounds of opening and closing doors are used to indicate that someone else has connected or disconnected from a user's video feed. They use stereotypical and unobtrusive sounds to make people aware of events in the workspace without interrupting normal workspace activities. We follow this approach attempting to discriminate in the more delicate domain of presence. ShareMon [13] used auditory icons to notify users about background file sharing events. To indicate the various actions involved, Cohen experimented with three types of sound mappings. For example, to indicate user login he used knock-knock-knock (iconic), "Kirk to enterprise" (metaphoric), and Ding-Dong (symbolic). To some degree, all three methods were intuitive and effective at communicating information, and users found them less disruptive than other modalities, like graphics and text-to-speech. In our work, we try to understand how these mappings affect users when they are used to interject ongoing communication.

The OutToLunch system [11] attempted to recreate an atmosphere of "group awareness". It gave isolated or dispersed group members the feeling that their coworkers were nearby, and also a sense of how busy they were, by taking advantage of the human ability to process background information using sound. Each user had a theme that was mixed in with a seamless loop of solo guitar music, and would only play when the user was typing on their keyboard. With only six people in the group, the paper reports that users had no trouble associating a theme with the person it represented. We attempted to use a similar approach, but when convolved with conversation, the multitrack instrument sounds over- load the channel and can be annoying. Similarly, there has been work in group- ware systems to address the issue of awareness through audio. GroupDesign, a real-time multi-user drawing tool, used audio echo to represent user action on another users' interface [26]. In Thunderwire [27], an audio-only shared media space, the audible click of a microphone being switched on or off served to let participants know when people were joining or leaving the discussion. In Chalk Sounds [14], Gutwin et al. used the granular synthesis method to create chalk sounds that were parameterized by the speed and pressure of an input stylus. Our goal is to extend such awareness without the need for users to break the flow of conversation by having to ask about who has joined or left the conference call, for instance.

3 AUDIO DESIGN FOR CAMEO

Synthesizing prior work on audio signals, we experiment with feedback techniques to support the social process of communication. Specifically, we incorporate them in the design of three CAMEO features — Speaker Identification, Audio Presence, and Entry/Exit announcements [15]. We focus on the functionality and design of assistive audio cues for each of these features.

Adding audio cues on to an already overloaded communication channel with multiple speakers provided ample opportunity to be distracting and inconsider- ate. We had to try many approaches and various kinds of audio feedback cues to find reasonable candidates to formally evaluate. We chose to work with Apple's GarageBand[1] software. We pre-recorded two ten-minute teleconference calls between five people and used these as the base tracks in Garageband. This was then overlaid with earcons and auditory icons, which included instrument sounds from http://free-loops.com/, and nature sounds from http://www.naturesoundsfor.me/. In the following paragraphs, we summarize the lessons learnt from the broad explorations in the designs of each of these three features.

3.1 Speaker Identity

One of the major reasons people might find video conferencing attractive is because it elevates identifying and distinguishing between speakers to a separate channel with less crosstalk [2, 6]. The question of identity and presence can get even more muddled when people on the line are not familiar with each other, or their accents. So we experimented with different audio cues to support speaker identification in such difficult situations.

We focussed on designing earcons that were easy to perceive, remember and discriminate. We initially experimented with assigning background instrumental tracks to each participant. For example, the bass track might be assigned to participant one, and the rhythm guitars to participant two. When the participants spoke, the track assigned to them would start to play in the background. This, however, proved to be distracting as the instrumental tracks introduced too much crosstalk on the channel.

To reduce crosstalk, we experimented with simple tones instead of tracks, that pulse while the participant speaks. We were able to achieve good discriminability by using the following timbres : tambourine, bongos, and vibraphone for the 2nd, 3rd and 4th participants, respectively. For the 1st and 5th participants we used muted electric bass with tones that were an octave apart.

Once these qualitative design evaluations were done, the next part of the audio design was to determine the temporal nature of the cues, i.e when they should play and for how long. The cues were designed to play at the beginning of every utterance a participant makes while holding the floor. We found that this worked best when the cues began playing a second into the utterance, as opposed to right at the beginning of the utterance. This duration was long enough to ensure that a participant was contributing more than just back-channel feedback, like uh-huh. With regards to the duration, a cue that was a second long was distracting, whereas a cue that was 0.25 sec long was too subtle to discriminate from the other cues. Thus, the duration of the cues was set to be 0.5 sec.

An interesting side effect from designing the audio cues in this manner was that they also seemed to emphasize a participants hold on the floor, reinforcing personal audio space. These can be thought of as analogous to people's use of

[1] `http://www.apple.com/ilife/garageband/`

hand gestures while speaking. Thus, when a speaker is speaking loudly and at a rapid pace, the cues pulse rapidly too. If the speaker is speaking softly and at a slower rate, the cues pulse at a slower rate.

The timbre from Garageband are high quality, and occupy a large portion of the soundscape when mixed-in with the conference call. To push the auditory cues to the background we experimented with a number of filters and reverb effects. We found that using a high-pass filter, and the “small room” reverb effect worked most effectively.

Another technique to aid with the identifying speakers is to spatialize them in a 2D environment. We used stereo panning to place the 5 speakers at the -32, -16, 0, +16, +32 positions on the Garageband’s Pan Dial. Like in experiments done by Ramloll [24], using stronger panning was actually distracting, and made the listener want to cock their head to the side where the sound was coming from, like when someone taps your shoulder.

Next, we describe the process of designing audio cues for the second CAMEO feature, i.e. to indicated presence of other collaborators on the line to a user.

3.2 Audio Presence

Feedback is very important for communication. Even the absence of feedback about whether the others on the line can hear you or not, can be distracting. For instance, without the addition of sidetone users have a tendency to attend to their displays to know if their call has been dropped or not. This can be dis- ruptive to the conversation. Sidetone is a form of feedback thats picked up from the mouthpiece and instantly introduce into the earpiece of the same handset. It gives users the assurance that their signal is being registered by the phone system, and is therefore now incorporated into most phone devices. Similarly, the awareness that the other participants are on the line, and are listening is important confirmation which reduces uncertainty about the channel continuing to be functional. There are various methods to present such information visually [6], but the visual channel comes at a high bandwidth and attention cost as described above.

Initial ideas involved the use of background sounds to create a soundscape around which users could orient themselves. For example, if Ron is playing music in the background, or Joe is driving, it becomes immediately obvious when either of them go offline, even if they aren’t talking. Their absence becomes conspicuous because of the sudden change in the soundscape. To test this idea in our system, we assigned different instrumental tracks to the participants, similar to the OutToLunch system [11]. The tracks would play while the user was on- line, but they overloaded the conversational channel and were, therefore, highly distracting.

To soften the crosstalk, we were inspired by radio shows where people call in to chat with the host. They are usually located in different audio environments, and it’s easy to tell them apart. We improvised on this by experimenting with different nature sounds like from a flowing stream, waves landing on the shore of a beach, and the chirping of birds. These blended so well into the background, though, that it was hard to notice when they stopped or started.

An alternative idea was to employ a roll cqall, i.e. to periodically announce the presence of the participants either by name or auditory icons. Announcing the names of participants would be the easiest to understand, but people may find it annoying to hear their names being called out periodically. Instead, auditory icons were created by sampling backchannel like participants laugh, and other characteristic sounds. These were inserted in the channel at periodic intervals when a participant had been silent for a while. However, they were too subtle and weren't noticed. We then experimented with recorded ambient environmental sounds of someone typing on a keyboard, clicking a mouse, opening and closing a drawer, and thumbing through papers. These auditory icons were found to be distinct, perceptible, and natural in a work environment.

In the final subsection, we consider multiple entry & exit announcements for CAMEO's Entry/Exit feature, which we describe below.

3.3 Entry/Exit

It can be hard to tell when participants get dropped from or reenter a conference call. Some existing conference call systems circumvent this by announcing entry and exit with a lengthy statement, which can be annoying. In this work we explore alternate iconic and metaphoric prompts to announce these events.

To avail of iconic mapping, we attempted to use the sounds of a door opening and closing to indicate entry and exit. We initially tried to superimpose the name of the participant with the sounds of the door opening and closing, but it was hard to discriminate between the door open and the door close sounds. We obviated this by appending the sounds to the name using different sequences. We usually hear the door open and then see the person enter, so an entrance is announced by the sound of a door opening followed by the name. On the other hand, when leaving we see a person head to the door and then hear the door close. So an exit is announced by the name followed by sound of the door closing.

We wanted to compare this to more metaphorical mapping approaches, like using fade and intonations. We modified the TTS in Audacity[2] by using the amplitude fade-in and fade-out effects for entrance and exit, respectively. But this reduced the understandability of the name. For intonations, we chose to map entrance to a normal intonation, and exit to an upward intonation. This was partly due to convenience as Apple's TTS engine automatically intonates a word when it is punctuated with a question mark. For instance, "Armstrong?" is automatically intoned upwards and indicates that a participant named Armstrong has been disconnected from the conference call.

In the next section, we formally evaluate a final set of audio designs and protocols that came out of the preliminary explorations. We conducted three studies that we discuss below for each of the three features.

[2] http://audacity.sourceforge.net/

4 STUDY I: Speaker Identification

The goal of this study is to understand how the addition of audio cues to a conference call can help people differentiate between different speakers. To highlight difficulties in recognizing people in a conference call we chose five non-native english speakers (males, 22-28 years old), and had them remotely collaborate on a sub-arctic airplane crash scenario commonly used in team building exercises. According to the scenario, the five of them had just survived a plane crash in Northern Canada, and had managed to salvage some items. Their task was to list the items in order of importance, and to come to an agreement on this order as a group. The audio from each participant was recorded in separate files, which was then processed in Audacity. Garageband was used to create three separate versions: a simple downmixed version; one mixed with the speaker identifica- tion earcons discussed in the previous section; and another which arranged the speakers spatially in a 2D environment. We compare these three and see how well participants do on each. Our hypothesis was that the participants will do better with the spatialization and earcons aids, than without any audio cues.

4.1 Methods

The speaker identity study asked participants to listen to a segment of a pre-recorded conference call while answering questions related to the conversation at hand, and speaker contributions.

Participants, Procedures, and Task. Thirty-two people were recruited for the study (8 female, 24 male). Participants were between 20 and 30 years old, and all reported having no hearing impairments. Participants had the choice to take part in the experiments either remotely or in the lab

This study had two stages, a training stage and a test stage. During the training stage, the participants were first asked to listen to the recorded introductions from the five speakers on the recorded conference call. They were presented with five colored buttons with the number and name of the different speakers. Upon clicking the button, they would hear a recording of the corresponding speaker saying their name and a fun fact. The next page allowed the participants to practice speaker tagging. They could click on the practice button which would cause the program to randomly play a short segment of the recorded conference call. The participant was asked to identify the speaker in the short segment by click- ing on the speaker's corresponding button. After every attempt both the right answer and the selected answer were displayed. The participant could choose to go to the next screen whenever they felt confident of successfully differentiat- ing between the different speakers. In the test stage, participants listened to a two minute and thirty second clip of the pre-recorded conference call. As they were listening, questions would appear about the conversation that had to be answered within five seconds.

Apparatus and Sounds. Participants were provided with Logitech headsets. Remote participants were requested to find a quiet place and use headsets. They were provided with the address of the server where the experiment was hosted, and were asked to access it using the Chrome browser.

The earcon audio cues were obtained from Garageband. Speakers one to five were assigned muted electric base (low tone), tambourine, bongos, vibraphone, and muted electric base (high tone), respectively. Similarly, for creating a spatial 2D environment, speakers one to five were placed at the -32, -16, 0, +16, and +32 units on Garageband's pan knob (2D spatial positioning).

During the training stage when participants were introduced to the five speakers, their corresponding instrument or spatial location was also displayed, both visually and aurally.

Study Design. We used a between group study, where half the participants answered the questions with the aid of musical instrument earcons, while the other half used 2D spatialization. For each group we also included a within- subject condition to compare the test condition (with earcons) to a baseline (with no earcons). To balance out learning effects, half the participants started with the baseline, while the other half started with the test condition. Different two-and-a-half minute segments of the pre-recorded conference call were used in the within-subject study. The first segment had nine questions, while the second segment had eight questions.

To keep the test conditions same across study participants, and to isolate only the participant's perception of the audio cues, we used the same pre-recorded tracks in our evaluations. A limitation of this approach is that the audio cues are evaluated by third-party observers, and not by active participants of a meeting. We tried to account for this by asking questions that were of a "who said <something related to conversation>" nature, which is different from asking who just spoke. The aim was two-fold. First, to keep the participants engaged in the conversation, and to prevent them from simply matching audio cues to the speaker. Second, to cognitively load the user (as they might be while participating in a conference call) so that the distractive effects of audio cues might come to bear on the results.

4.2 Results

We present our results below in terms of participants being able to accurately identify the speakers on a conference call, and their response times. Attempt rate is the fraction of questions users answered in each condition. For accuracy, we report two metrics: Overall Accuracy, which includes questions that were not answered, and Attempt Accuracy, which only includes questions that were answered. Together, these metrics should account for distractions that audio cues might introduce causing participants to take longer than five seconds to answer a question.

Spatialization vs. No audio cues. Participants ability to identify speak- ers increased significantly, with greater than 20% improvement using spatial audio cues. They were able to do this almost half a second quicker on aver- age when compared to the condition without audio cues ($p<0.05$, 1-tailed t- test, Table 1). Overall Accuracy: SEM=(0.024, 0.041); Attempted Accuracy: SEM=(0.046, 0.044); Response Time: SEM=(151.2, 163.7); N=16.

Table 1. Accuracy metrics and average response times for speaker identification with and without 2D spatialization

	Accuracy (Overall)	Attempt Rate	Accuracy (Attempted)	Response Time (ms)
Spatial	0.573	0.841	0.691	2187.4
No cues	0.435	0.788	0.570	2663.0

Earcons vs. No audio cues. With earcons, participants were also able to achieve an increase in accuracy of 30% on average over the condition with no audio cues ($p<0.05$). Participants also appeared quick to respond but the difference was not significant ($p<0.1$, 1-tailed t-test, Table 2). Overall Accu- racy: SEM=(0.052, 0.024); Attempted Accuracy: SEM=(0.054, 0.054); Response Time: SEM=(169.0, 111.5); N=16.

Table 2. Accuracy metrics and average response times for speaker identification with and without earcons

	Accuracy (Overall)	Attempt Rate	Accuracy (Attempted)	Response Time (ms)
Earcons	0.538	0.772	0.703	2257.8
No cues	0.386	0.819	0.475	2537.2

Summary. We were able to show that speaker identification improved with the addition of either spatial cues, or earcons. A between group analysis did not reveal any difference between these two conditions. Furthermore, there was no significant difference in the number of questions that were attempted across the three conditions from which we might infer that the addition of audio cues was not notably distractive.

5 STUDY II: Audio Presence

The goal of the audio presence study is to investigate whether the addition of audio cues to a conference call can help reassure people that the other partici- pants are still on the line, and haven't been disconnected. A different segment of the pre-recorded conference call described above was used in this study. Garage-band was used to create two separate versions: a simple downmixed version with no audio cues added; and one mixed with the auditory icons for audio presence discussed in the previous section.

5.1 Methods

Participants were asked to listen to a segment of a pre-recorded conference call while answering some questions related to the conversation. The participants were also asked to indicate if they thought a participant had been dropped from the call.

Participants, Procedures, and Task. Twenty people were recruited for the study (4 female, 16 male). Participants were between 20 and 30 years old, and all reported having no hearing impairments. Participants had the choice to take part in the experiments either remotely or in the lab

The participant was asked to listen to a five-minute clip of the pre-recorded conference call. As they were listening, questions would appear about the con-versation that had to be answered within ten seconds. Participants were also instructed to periodically ensure that everyone was online. They could do so by pressing the "nudge" button which simulated feedback from each participant stating that they were still there (like a ping test).

Apparatus and Sounds. The apparatus used by the participants is identical to the first study. The cues in this study were recorded using an iPhone, and processed in Audacity. As motivated in our exploration experiments above, these include auditory icons of ambient environmental sounds like someone typing on a keyboard, clicking a mouse, opening and closing a desk drawer, and shuffling through papers. These cues were then added to the segment of the pre-recorded conference call used for this test.

Study Design. Our working hypothesis was that adding audio cues like key- board sounds and mouse clicks acted to reinforce the presence of people who had not spoken in a while, but were still online. In other words, we wanted to show that like the sidetone, adding audio cues improves awareness about the presence of other collaborators.

We used a between group study, where half the participants were presented with audio cues (test condition), and the other half was not (baseline condition). Participants had to answer eight multiple-choice questions while listening to the conversation. This was to simulate a real meeting where participants would be paying attention to the conversation, and not actively tracking the presence of other collaborators. Participants were told that because of some collaborators being in weak signal areas, there was a high chance that they might accidentally drop off the call. They were asked to virtually "nudge" the other participants if they suspected that one of them was not present.

5.2 Results

We investigate our hypothesis by comparing how often users "nudge" others to check if they are present, with and without the auditory icons discussed above. We found that the number of nudges was reduced by 37% in the condition where the auditory icons were used ($p<0.05$, 2-tailed t-test, Table 3). There was no significant difference in the attempt rate or error rate. # of Nudges: SEM=(0.72, 0.64); N=10.

Table 3. Average number of nudges, attempt rate, and error rate with and without auditory icons

	# of Nudges	Attempt Rate	Error Rate
Auditory icons	3.50	0.90	0.33
No audio cues	5.63	0.81	0.35

6 Study III: Entry and Exit

The goal of this study is to understand the effects that different conference call entry & exit announcements have on the participants, and meetings in general. We focus on three kinds of prompts, namely, speech, iconic and metaphoric. Our hypothesis is that the metaphoric prompts using different intonations will have the least impact on participants cognitive capability (i.e., their ability to follow game protocol in this particular study).

6.1 Methods

To bring out the effects, we designed and built a memory card game for four people that can be accessed remotely from the browser. Participants are paired off into two teams that take turns in choosing two cards from the sixteen that are shown face down on a GUI screen. If the two cards chosen by a team match, the team wins the turn. The team that matches the most number of pairs, wins the game.

Table 4. Entry & Exit prompts using different mappings in each test condition

	Entry & Exit Prompts
Speech	<participant_name> has joined the conference <participant_name> has left the conference
Iconic	sound of door opening + <participant_name> <participant_name> + sound of door closing
Metaphoric	<participant_name> (said with normal intonation) <participant_name> (said with raising intonation)

Participants, Procedures, Task. We recruited 21 participants for this study (4 female, 17 male). Participants were between 20 and 30 years old, and collaborated remotely on the game. Six unique groups of four participants each were tested (some participants repeated).

When the participants join the meeting, the administrator would introduce them to the game, and the protocol they were to follow. During a team's turn, both team members are required to select a card. The selected card is revealed only to its selector. Thus, the first team member to click open a card has to communicate its content and position, based on which their partner picks the second card. The protocol specifically requires the team partners to alternate who gets to pick the first card at every turn. The protocol was designed in this way to encourage discussion.

After a practice round, the administrator would notify the participants that the experiments were going to begin. They were told that during the experiments, participants would randomly be dropped from the meeting. If they happened to be dropped from the conference call, they were requested to rejoin as soon as possible. During the course of such an event, a prompt would play to notify the rest of the participants that someone had left the conference call, while another prompt would play to indicate that they had joined back.

Apparatus and Sounds. The apparatus used was identical to the first two studies. Mumble[3] was used to host the conference call. All the participants were requested to download the Mumble client and follow the instructions that were provided.

A mac mini was used to run the python script that generated the prompts. The three entry and exit prompts that were used are speech-based, iconic, and metaphoric (Table 4). The prompts are dynamically created using Apple's text-to-speech engine, and pre-recorded audio of a door opening and closing. The Python script was also set up to use the Mumble server's Ice remote procedure call interface to arbitrarily disconnect people every thirty seconds.

[3] `http://mumble.sourceforge.net/`

Study Design. We used a within-subject study where each group played three rounds of the memory game, one for each of the three conditions. To balance out any learning effects, different sequences of the conditions were used for each group (Table 4).

6.2 Results

We wanted to investigate the effect that the different prompts would have on the participants ability to observe protocol, i.e. team members switching turns to pick the first card. We only take into account turns where both participants are online. We found that the metaphoric prompts had the lowest error rate at

15% in participants ability to maintain protocol compared to both the iconic and speech prompts ($p<0.05$, 2-tailed t-test, Table 5). The iconic prompts affected the participants as badly as the speech prompts did with error rates larger than 25%. Error Rate: SEM=(0.05, 0.05, 0.05); N=6.

Table 5. Average error rates in following the protocol and game duration across the three conditions

	Error Rate	Duration (sec)
Speech	0.29	262.3
Iconic	0.26	258.6
Metaphoric	0.15	222.0

We also wanted to understand how the different prompts affected the game. We hypothesized that the shorter prompts would create less disruptions allowing the participants to finish the game quicker. There wasn't a significant difference in the durations, but the participants do appear to finish the games faster in the condition with the metaphoric prompts. The average durations are shown in Table 5. Duration: SEM=(34.4, 22.3, 16.3); N=6.

Participant Preferences. During the pilot experiments, participants strongly preferred the speech prompts to the metaphorical ones, which they found to be ambiguous. They were largely ambivalent about the iconic prompts. To help disambiguate the prompts in general, we began playing each of them at the start of their respective test conditions. This practice saw an increase in the number of participants who preferred the metaphoric prompts as they found it to be less distracting. They remained neutral with regards to the iconic prompts, although some of them claimed that it was hard to distinguish between the door opening and closing sounds when the line was noisy. This might explain poor participant performance under the iconic condition.

7 Discussion

The three studies show that a constricted audio communication channel can be augmented with assistive social feedback cues, even in highly dynamic environments. In specific, we empirically showed that these cues allowed users to identify speakers more accurately, increased awareness about the presence of other collaborators, and improved participant performance.

To demonstrate the utility of the audio cues, we simulated particularly dif- ficult and stressing situations. It is hard for the average person to distinguish between five people of the same gender with similar accents, or notice the quiet person in the room when engaged in conversation, especially when they aren't visible. Similarly, keeping track of multiple things while coordinating with others is difficult when there are a lot of distractions in the environment. As interactive systems weave themselves tighter into our social fabric, they need to be designed so as to accommodate such typically complex social situations. For instance, the first time we are introduced to a team that we are collaborating with, is when our understanding of their speech is most important; but it is also when their accents and behaviors are most difficult to interpret. We are very excited about being able to add audio to an already loaded channel and improve the perceived and real understanding of the situation.

Similarly, in an increasingly mobile and global workforce, a user might be in a noisy environment and have trouble distinguishing between some of the other collaborators on the conference call. In this case, the user could choose to add cues to some of the other collaborators, which would play only on their own channel. The sounds might also act as aids to users who might choose to associate meta-information (like the person's location or function) with an audio cue. Likewise, when to use speech, iconic or metaphoric prompts to announce events might be dependent on the situation. Developing an understanding of how these cues affect participants, and their applicability in different situations, allows us to build a vocabulary of actuators that a considerate agent like CAMEO would know when and how to use.

8 Conclusions and Future Work

The work described here focuses on domain-independent social feedback. We show how careful choice of its syntax — its sound and placement (like long utterances), and semantics — its direct and indirect relationship to the conversational channel (such as putting speech on top of a conversational channel), can deeply affect its goal of supporting social interactions. In particular, this paper focuses on how audio cues like earcons and auditory icons can appropriately provide feedback to stymie the disorienting effects of technology mediation. We show that earcons can improve accuracy on speaker identification, and is comparable to 2D spatialization of speakers. Auditory icons, like keyboard typing and mouse clicking, can act as feedback to reassure participants about the presence of others on the line. We also show that using metaphorical prompts (intonations) to announce events like entry and exit of participants reduces errors when compared to speech prompts.

With the proliferation of phones and tablets, more and more interactive systems are being used in social settings. The imperative now is for technology that celebrates situational awareness, and appropriateness. While this work builds a vocabulary of effectors to improve teleconference meetings, we look forward to the opportunities afforded by developing an understanding of how to accom- modate system feedback in variety of other social situations. The choices of when and how a considerate agent should intrude on a communication channel is shown here to be delicate but tractable. We are excited about the possibility in the utility of such considerate agents across other interactive scenarios that would benefit from a system's ability to regulate and coordinate social feedback.

References

1. Higgins, E.T.: Achieving 'shared reality' in the communication game: A social action that creates meaning. Language and Social Psychology 11(3), 107–131 (1992)
2. Sellen, A.J.: Remote conversations: the effects of mediating talk with technology. HCI 10(4), 401–444 (1995)
3. Halbe, D.: Whos there? Business Communication 49(1), 48–73 (2012)
4. Brubaker, J.R., Venolia, G., Tang, J.C.: Focusing on shared experiences: moving beyond the camera in video communication. In: Proceedings of the Designing Interactive Systems Conference, DIS 2012, pp. 96–105. ACM (2012)
5. Nguyen, D.T., Canny, J.: Multiview: improving trust in group video conferencing through spatial faithfulness. In: Proc. CHI 2007, pp. 1465–1474. ACM (2007)
6. Yankelovich, N., Walker, W., Roberts, P., Wessler, M., Kaplan, J., Provino, J.: Meeting central: making distributed meetings more effective. In: Proc. CSCW 2004, pp. 419–428. ACM (2004)
7. Tang, J.C., Isaacs, E.: Why do users like video? In: Proc. CSCW 1992, vol. 1(3), pp. 163–196 (1992)
8. Selker, T.: Understanding considerate systems – UCS (pronounced: You see us). In: 2010 International Symposium on Collaborative Technologies and Systems, pp. 1–12. IEEE (2010)
9. Arons, B.: A Review of The Cocktail Party Effect. The American Voice I/O Society 12, 35–50 (1992)
10. Gaver, W.W.: Sound support for collaboration. In: Proc ECSCW 1991, pp. 293–308. Kluwer Academic Publishers (1991)
11. Cohen, J.: Out to lunch: Further adventures monitoring background activity. In: Proc. ICAD 1994, Santa Fe Institute, pp. 15–20 (1994)
12. Rigas, D.I., Hopwood, D., Memery, D.: Communicating spatial information via a multimedia-auditory interface. In: Proc. EUROMICRO 1999, vol. 2, pp. 398–405. IEEE Computer Society (1999)
13. Cohen, J.: "Kirk here": using genre sounds to monitor background activity. In: Proc CHI 1993, pp. 63–64. ACM (1993)

14. Gutwin, C., Schneider, O., Xiao, R., Brewster, S.: Chalk sounds: the effects of dynamic synthesized audio on workspace awareness in distributed groupware. In: Proc. CSCW 2011, pp. 85–94. ACM (2011)
15. Rajan, R., Chen, C., Selker, T.: Considerate Audio MEdiating Oracle (CAMEO): improving human-to-human communications in conference calls. In: Proc. DIS 2012, pp. 86–95. ACM (2012)
16. Edwards, A.D.N.: Soundtrack: an auditory interface for blind users. HCI 4(1), 45–66 (1989)
17. Gaver, W.W.: The SonicFinder: An Interface That Uses Auditory Icons. HCI 4(1), 67–94 (1989)
18. Strachan, S., Eslambolchilar, P., Murray-Smith, R., Hughes, S., O'Modhrain, S.: GpsTunes: controlling navigation via audio feedback. In: Proc. MobileHCI 2005, pp. 275–278. ACM (2005)
19. Schlienger, C., Conversy, S., Chatty, S., Anquetil, M., Mertz, C.: Improving users' comprehension of changes with animation and sound: An empirical assessment. In: Baranauskas, C., Abascal, J., Barbosa, S.D.J. (eds.) INTERACT 2007. LNCS, vol. 4662, pp. 207–220. Springer, Heidelberg (2007)
20. Dingler, T., Brewster, S.: AudioFeeds: a mobile auditory application for monitoring online activities. In: Proc. MM 2010, pp. 1067–1070. ACM (2010)
21. McGookin, D., Brewster, S.: PULSE: the design and evaluation of an auditory display to provide a social vibe. In: Proc. CHI 2012, pp. 1263–1272. ACM (2012)
22. Gaver, W.W., Smith, R.B.: Auditory icons in large-scale collaborative environments. In: Proc. INTERACT 1990, pp. 735–740. North-Holland (1990)
23. Gaver, W.W., Smith, R.B., O'Shea, T.: Effective sounds in complex systems: the ARKOLA simulation. In: Proc. CHI 1991, pp. 85–90. ACM (1991)
24. Ramloll, R., Mariani, J.: Do localised auditory cues in group drawing environments matter? In: Proc. ICAD 1998, p. 24. British Computer Society (1998)
25. McGookin, D., Brewster, S.: An initial investigation into non-visual computer supported collaboration. In: Ext. Abstracts CHI 2007, CHI EA 2007, pp. 2573–2578. ACM (2007)
26. Beaudouin-Lafon, M., Karsenty, A.: Transparency and awareness in a real-time groupware system. In: Proc. UIST 1992, pp. 171–180. ACM (1992)
27. Hindus, D., Ackerman, M.S., Mainwaring, S., Starr, B.: Thunderwire: a field study of an audio-only media space. In: Proc. CSCW 1996, pp. 238–247. ACM (1996)

Supporting Voice Content Sharing among Underprivileged People in Urban India

Christian Remy[1], Sheetal K. Agarwal[2], Arun Kumar[2], and Saurabh Srivastava[2]

[1] University of Zurich, Binzmühlestrasse 14, 8050 Zurich, Switzerland
remy@ifi.uzh.ch
[2] IBM Research-India, Plot 4, Block C, Vasant Kunj, Phase II, New Delhi, India
{sheetaga,kkarun,saurabhsrivastava}@in.ibm.com

Abstract. Recent advances in voice-based telecom information systems enable underprivileged and low-literacy users to access and offer online services without expensive devices or specialized technical knowledge. We propose SRLs (speech resource locators), a mechanism that facilitates the creation, access, and sharing of online voice content. To test the interaction with SRLs, we developed a proof-of-concept application that allows for simple sharing of voice content. We subsequently created a smartphone application for the same service that provided a graphical user interface to the online voice application. Our findings show that literate underprivileged people were able to share online voice content on feature phones and smart phones whereas in low-literacy people were unable to access shared content over feature phones but able to do so on smart phones. We conclude by highlighting opportunities and challenges for the design of voice-based applications that support information sharing.

Keywords: HCI4D, Information Sharing, ICTD, User-Centered Design, Interactive Voice Systems, Smartphones, India.

1 Introduction

In recent years, a variety of voice-based telecom information systems have been created, allowing people in developing regions to access information they would not be able to retrieve otherwise. Among the many barriers to access a few stand out: low computer ownership, low levels of literacy, and lack of Internet access [8]. Interactive voice systems are services that users can call using easily accessible and affordable commercial mobile phones and navigate by speech or dual-tone multi-frequency (DTMF) input in order to retrieve information. Voice-based systems have been proven to be useful and accessible, through research projects [1, 19, 23] as well as large-scale deployments for public use [4, 13, 26]. Examples where these systems have been put into practice include but are not limited to: information for farmers in rural areas [19], education for school children [17], and job opportunities in urban

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 489–506, 2013.

areas [26]. One of the aspects that most of these applications have in common is the concept of knowledge distribution beyond face-to-face communication. An important issue when developing such applications is the sharing aspect, e.g., if a user wants to notify a friend about specific voice content she came across. We derived the following scenario that shows the need for sharing and the interaction breakdowns in current voice-based systems:

Raj is a migrant from a small village in Punjab who works part-time at a big market in Delhi. He comes across an interactive voice application about health information and posts a query on behalf of his cousin, who has recently been diagnosed with a new skin disease. A doctor records a response to the query, informing Raj of a new drug along with its usage instructions. Raj believes that this could help to improve his cousin's condition. Due to the complicated drug name and usage description, Raj is hesitant to call his cousin and paraphrase the information. Instead he would prefer to point him to the specific information.

There are several barriers illustrated in this example which are typical of voice-based telecom information systems. Currently, such applications do not offer easy ways to access or forward a specific piece of content; simply telling another person about the service results in a difficult navigation task to get to the information. Furthermore, everyday interactions with text content such as searching, browsing, or navigating voice content are still unsolved problems. In this particular example, direct access to a specific piece of voice content is the biggest challenge; this becomes even more complicated for information retrieval of dynamic content, such as user-generated content.

Further, given that pragmatic software for natural speech recognition is not yet available for most languages prevalent in developing countries, searching for audio content is in a nascent stage. Therefore, alternate mechanisms such as sharing content explicitly assume much more importance in the context of voice applications.

In analogy to URLs (uniform resource locators) for accessing Internet content, we therefore propose *SRLs* (speech resource locators) for voice-based systems. An SRL enables users to point to a specific item of voice content in an interactive voice application, thus enabling explicit sharing of online voice content. An SRL consists of a phone number (the voice-based application) and a *LinkCode* (a unique identifier for the online voice content). SRLs are novel as they offer users a way to create persistent links to content, dynamically while interacting with the site. The links are persistent as they can be shared with others even after the call has ended. This ability to link and provide direct access to content is so valued that startup companies[1,2] have created a business out of it. These companies essentially offer 'links' to a service (for instance, extension of a particular department or employee) desired by a caller. Instead of calling the enterprise Interactive Voice Response System (IVR), the caller calls one of these services which automatically traverses the touch tone hierarchy of the target IVR and connects the user directly to the desired option. SRLs are meant to fill the

---

[1] `http://deepdial.com/`
[2] `http://gethuman.com/`

gap in existing techniques by enabling a mechanism for sharing specific online voice content with a set of intended recipients. Furthermore, they enable it both for pre-created and user generated content. The links created by users themselves are usable across applications and users.

In order to test the accessibility, usability, and usefulness of SRLs, we developed a proof-of-concept application and conducted field studies in urban areas of two states in India. We highlight insights from the observations and the informal interviews during the user tests. Based on the findings from the first study, we conducted a second study, in which we evaluated the same interactive voice application with a graphical user interface augmenting the interaction with a smartphone application. Our findings suggest opportunities for the design of new interactive voice applications that allow content sharing, but also challenges that need to be addressed in the future.

We propose that SRLs can simplify the interaction required for knowledge distribution of voice content. Furthermore, we present a study that explores sharing of online voice content through low end mobiles and smartphones. The results of our studies provide empirical evidence that sharing of voice content can be supported by using SRLs. The findings highlight that smartphone applications can improve the user experience for sharing of voice content, in particular for low-literacy populations that were unable to receive SRLs on feature phones.

2 Related Work

By enabling the underprivileged to share voice content, we are providing alternative mechanisms for information retrieval and knowledge distribution among this population. Our background research focuses on two main aspects in previous works: firstly, to what extent sharing of information has been investigated in emerging and developing countries; secondly, what other approaches exist that could lend themselves helpful to achieve sharing and better information distribution for the underprivileged.

2.1 Content Sharing for the Underprivileged

Smyth et al. [25] investigated media sharing on mobile phones and discovered that people in urban India adapt quickly to new interaction techniques if the application addresses their needs while saving them money at the same time. Particularly, if the motivation is high enough, users will overcome smaller interaction issues and find a way to use it even in their everyday practice. However, Densmore [11] pointed out that overly high barriers for sharing result in failure, as in their case relying on Internet connectivity for smartphone applications in a project in Uganda.

Dhir et al. [12] conducted ethnographic studies about information sharing in South Africa, observing that face-to-face communication was by far the most commonly used communication channel. However, especially young people or those involved in businesses express a strong desire for new technology.

Polly [21] is a service designed to facilitate communication among underprivileged users by implementing a service that calls users back, reducing telephony costs significantly. Another system that has been implemented and evaluated in depth is Message Phone [15], which focuses on exploring the usefulness of voice messages for underprivileged people in Uganda by comparing it to SMS messages. Having similar functionality, Bubbly[3] and Kirusa[4] are two systems that deploy fully functional solutions for voice messaging in developing countries. The proliferation of such voice-based services yields strong evidence for the need of alternative means of sharing, as also pointed out in the introduction.

2.2 Improving Voice-Based Information Accessibility

Dhanesha et al. [10] demonstrated bookmarking content in voice-based applications for users and how to re-visit that content. However, sharing of bookmarks is not available. Bookmarking entails assigning an identifier, which must be unique for a caller. The application session information required to reach that point in the application is saved in order to revisit the bookmark. Bookmarking audio content enables the caller to retrieve content of interest to him quickly when he calls the voice application. However bookmarks are typically not shareable and are valid only for the user who created them.

A different concept is to augment existing websites with voice navigation, such as TeleWeb [7], Hearsay [20], or the introduction of voice anchors [22]. This voice anchors allow users to navigate HTML websites with a feature phone by converting information automatically into VoiceXML. By placing a voice anchor via voice input, the user can then go back to the auditory representation of the website later on. Furthermore, the text-to-speech output does not offer the same user experience as voice content recorded by a real person.

To provide search within voice-based systems, Ajmera et al. [2] introduced algorithms that allow for automatic tagging of audio documents in voice-based information systems. Similarly, Srivastava et al. [24] proposed SWAicons that assign auditory cues to improve navigation in voice-based information systems.

These approaches facilitate navigation in voice applications, allow for easier accessibility, and address problems of information retrieval. We present a different concept that enables sharing of any online voice content. While it addresses similar issues as previous projects, it does so by enabling not only direct access for one user to static content, but also allows users to point others to particular items of online voice content, including static and user-generated content.

3 The Billi Mausi Voice Application

We developed a voice-based telephony system that allowed for simple tasks focusing on the sharing mechanics. Our main goal was to implement a system that served as a vehicle to see whether SRLs are useful in lowering barriers to interacting with and

[3] `http://bubblemotion.com`
[4] `http://www.kirusa.com`

sharing voice content. We therefore opted for a single-service application that was easy to understand and not tied to any specific use case. In particular, the application let us focus on the three important parts of the sharing aspect: sending, receiving, and accessing SRLs through SMS, without heavy navigation or content-wise explanation in the voice application.

Furthermore, we wanted to attract a wide range of possible users, and therefore chose to use an application that would appeal to many people and was not constrained to a particular domain such as healthcare or education. To serve this purpose, we developed a voice application called *Billi Mausi*[5]. It is a playful service that makes use of the character of a talking cat as the anchor for the voice application[6]. The Billi Mausi character prompts the caller to record her voice, which it then plays back with a modulated pitch (meant to represent a cat's voice). Recent studies conducted by HCI researchers also have shown that fun applications lend themselves well to designing for underprivileged users and encouraging them to participate in on-the-street user tests [17, 25].

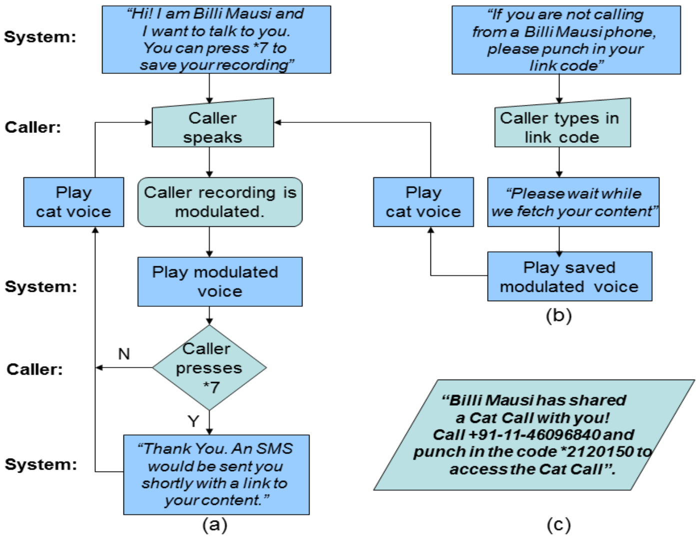


Fig. 1. Interaction tree of sharing voice site content, a) calling the general application and creating a recording b) calling the application to retrieve stored content, c) example SMS as received by our users

As shown in Figure 1 (a), on receiving a call the Billi Mausi application plays a short introduction (*"Hi, my name is Billi Mausi and I want to talk with you!"* in Hindi). It also informs the caller that by pressing a key combination (*7) on the phone's dial pad, a recording can be stored, and plays a "meow" cat sound as soon as it is

---

5 Billi Mausi (Hindi for *cat aunt*) is a well-known fictional representation of cat as a character in many children's stories and poems in India.

6 A popular service with a similar functionality is Talking Tom, an application available for iPhone and Android.

ready to record user input. Once the input is completed, a different cat sound is played while the server processes the input (which usually takes about a second) and then plays back the modulated voice to the user.

If the caller presses *7 while the modulated voice is played back, or within a few seconds after the playback (we had it configured to three seconds), the voice application (an Apache Tomcat server) registers that as a request to save the recording. At this point, the voice application logic stores the recording, generates an SRL for it, and announces the scheduled delivery of an SMS to the caller. These SRLs allow users to bypass cumbersome and time-consuming voice menu navigation and connect directly to the desired content in an online voice application. An SRL is a pair consisting of a phone number followed by a numeric code (called *LinkCode*) preceded by an asterisk ('*') that can be entered on the voice application as a dual-tone multi-frequency (DTMF) signal. An example SRL would be <+911146096840, *21250> as shown in the sample SMS in Figure 1 (c), that is sent to the caller on successful saving of recording and generation of SRL.

As shown in Figure 1 (b), an SRL can be accessed by simply calling the specified phone number and entering the LinkCode on the phone's keypad, once connected. If the caller is calling from a 'Billi Mausi phone', i.e., an Android smartphone with the Billi Mausi Android application installed, the manual step of punching the LinkCode is not required. Once the LinkCode has been received by the voice application, it fetches and plays the corresponding content and continues with the regular Billi Mausi voice application interaction (i.e., the caller can record and share more such recordings). Thus SRLs not only point to online voice content, they enable callers to continue interacting with the application from that point.

4 Study #1: Sharing with Feature Phones

The user studies were conducted in urban areas of two states of Northern India, namely Uttar Pradesh and New Delhi. At the onset user tests were conducted in the field as well as in the laboratory. Preliminary observations showed that feedback in the lab setting was not as helpful or as successful in eliciting honest participant feedback as field studies. Similar issues have been observed previously in research projects in India [5, 3]. We therefore decided to conduct the remaining user tests entirely in the field to get more realistic feedback and let participants use the system in their everyday environment.

4.1 User Test Protocol

The user test started with an introduction into what our service does, with an emphasis on the sharing part, but building on the funny interaction with Billi Mausi's cat voice to keep participants engaged. We provided them with a short demo of the system and asked them to use it themselves, with a task consisting of four different steps: 1) call the Billi Mausi application, 2) record a sentence and store it by pressing *7, 3) read the incoming SMS, and 4) follow the instructions in the SMS to access the recording. The tasks were intentionally kept simple to focus on the sharing mechanics only and not be overshadowed by other usability issues or content-related questions.

We closely observed the participants while they used the system and took notes of all interaction breakdowns as well as any comments they made. Once the participant completed the task and listened to the recording, we conducted a short interview, asking for demographic background questions (age, occupation, technical background, as in phone model and Internet experience, if any) as well as feedback on the system. The debriefing interview had two separate goals: firstly, we wanted to ensure that the participants understood the sharing concept by asking clarifying questions and/or letting them explain in their own words what the system does. Secondly, we asked questions designed to elicit: 1) whether they liked the system, 2) if they would use the system and if so what for, 3) what ideas they might have for potential uses, and 4) what overall comments they had.

All user tests were conducted in the participants' native language (Hindi) by researchers who were fully proficient in the area's specific local dialect. There were at least two researchers for each user test: one to conduct the interview and one to observe the interaction and take notes. An average user test took about 20-30 minutes, in three parts of roughly equal length: the first part was about sparking participants' interest, explaining the system, and showing the demo; the second part involved participants completing the tasks; the third part concluded the user test with the debriefing interview.

4.2 User Population

We conducted the studies in urban areas, where we recruited participants from malls, street markets, residential area and their vicinities, and taxi stands. The user set primarily consisted of migrant workers from the unorganized sector, which accounts for 90% of India's workforce [18].

Our study comprised 38 users, of which 33 were male and five female. Participants' ages ranged from 19 to 55 (average 31.2). The five female participants were housemaids in the communities we conducted our studies in, but the atmosphere of these user tests was more comparable to a lab study than to a field study. A group of men approaching a female stranger is not considered culturally appropriate in India and the housemaids acted almost intimidated during the user tests. Therefore, we decided against recruiting further female participants. Our observations in this regard parallel previous experiences with the difficulty with interviewing female users in underprivileged communities (e.g., [3]). Thus, the user population in our studies comprised occupations such as several kinds of drivers (auto rickshaw, cab, and bus), plumbers, carpenters, shop keepers, cleaning staff, and security guards in addition to the aforementioned housemaids. For most of these occupations, the gender distribution is close to 100% male in Delhi.

5 Results

We categorized participants into three rough levels of literacy: illiterate, low-literacy, and literate. In order to estimate their literacy, we asked them to read and paraphrase the SMS that they received during the interview. We also asked for their level of literacy, but often discovered mismatches in which people underestimated or overestimated their ability to read and understand the SMS. In our study, we refer to participants who could read some words and had a basic understanding of written language, but were unable to comprehend the SMS as a whole as "low-literacy." Out of 38 participants, four were illiterate and seven low-literacy (29% of all participants). We acknowledge that this is not a scientific approach to measuring literacy. However, it provided us with useful, albeit approximate, context about our participants without requiring more rigorous and extensive literacy testing that would have made their study participation inconvenient and burdensome.

5.1 Performance Results

Figure 2b shows the participants' ability to use the system including the SMS reception part, Figure 2a the level of understanding that participants had about the system and its sharing functionality, by level of literacy. The category "didn't complete" encompasses two different reasons: firstly, some participants were not able to use the system – even with our help. Secondly, some participants declined to use the system (e.g., because they felt it was beyond their abilities) after they had patiently listened to the introduction and looked at the demo. They read and correctly paraphrased the SMS and are therefore regarded as literate in our study, but did not complete the user test; e.g., one such participant said: *"I don't want to use it, but you can ask me questions."* Since these users answered all our questions, expressed interest, and provided insightful comments, we decided to continue the interview with them and did not exclude them from the study.

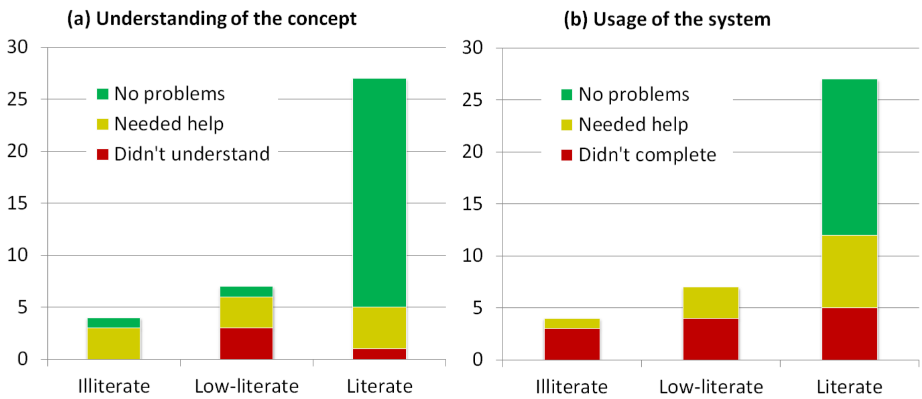


Fig. 2. Participants' understanding of the concept and ability to use the system

Low-literacy and illiterate participants had significant problems in using the system and understanding its use. Seven of the low-literacy or illiterate participants could not complete the last task (accessing the recording by following the instructions in the SMS) or did not even try since they could not read the SMS. This difficulty was also reflected in the interviews when participants described their issues with the system, which we elaborate on in the next section. Altogether, 26 out of 38 users (68%) completed all tasks, from creating a recording through following the instructions in the SMS to accessing the online voice content; 11 of them with help of the researchers.

In the interviews, participants revealed a much better understanding of the system than their use would suggest. Overall, 24 out of 38 participants were able to correctly paraphrase what the system did and proposed a simple use case that clearly indicated that they understood the sharing concept. Even the majority of low-literacy and illiterate participants were able to abstract the proposed idea, which was probably due to the more elaborate explanation and heavier assistance in using the application, giving them additional hints and examples. This is reflected by the category labeled as "needed help" to understand the system in Figure 2a.

5.2 Usability Issues

We received mixed feedback in our interviews. Although participants liked the system in general and only a few saw no use in it, the interaction was questioned by many. The following comment expresses a frequent response: *"This is a nice idea, but it's so difficult to use… I need a piece of paper to write down the LinkCode, call the number and then enter the LinkCode number… this is too complicated." (Shopkeeper, 26)*

These concerns, however, addressed accessing the voice content and not the sharing concept itself. While the idea of recording voice content was well received, the need to enter the LinkCode was criticized. Our participants had several ideas on how to overcome this issue, e.g.: *"Just let me enter the phone number of a person I want to share this with within the voice application directly, without this SMS." (Auto rickshaw driver, 38)*

Even though this solution is feasible, it breaks down when a user wants to share the content with multiple people. In this case the user must enter the number of each recipient, which not only results in an increased memory load on the user, but it might also be a more expensive solution. Moreover, the probability of entering the numbers incorrectly remains high and there is no way for the user to know if the content was indeed shared with all recipients. Our system allows storing the link to the content and sharing it at a later point in time; this would not be possible with an immediate sharing solution.

5.3 Use Cases and Opportunities for Sharing Content

Despite these issues, participants still saw a number of potential opportunities for them to use the application. Due to the playful context of our user test and the humorous voice in which our application stored and played back the caller's voice, some suggested sharing fun content: *"I'd use it to send jokes to my friends." (Driver, 31)*

Many participants said they mainly use their phone for business purposes, to talk to their employers or to their family. However, the sharing functionality combined with the presented application made them aware of new ways and uses of communication. One participant, a property dealer with a stand in the middle of the street on the outskirts of Delhi, suggested using the application for his business: *"Many clients receive SMSs with information about new available offerings. Using this application, I would send them a recording of my voice, which is much better than SMS, as voice is more personal; but I would only have to record the information once and I could share it with multiple people." (Property dealer, 27)*

This use case would require recording the message in a non-modulated fashion, as it is about serious content rather than entertainment. In fact, many suggested uses were about storing non-humorous content and dealing with real-world problems. Participants not only mentioned that they would use this to store recordings themselves, but also that they would like to receive recordings by others, e.g.: *"If I'm not available, my friends can record something and send me the link to this recording, and I can listen to it later." (Auto rickshaw driver, 30)*

Of course, a service like this suggestion already exists: voicemail. However, it is important to note that many of the participants who described these use cases to us had never heard of voicemail and did not use it. Furthermore, while voicemail is a one-to-one service, online voice content sharing is a one-to-many phenomenon; it enables more than just saving one message for one specific person but aims to provide means of sharing different voice content (user-generated and pre-existing information) with multiple people. Therefore, it goes far beyond what traditional voicemail services can offer and requires a different user interaction.

6 Conclusions of the First Study

The basic idea of SRLs was perceived very well; many participants expressed enthusiasm while using our system, which is in line with the positive feedback. However, low-literacy and illiterate users had difficulties or were unable to access the recording as they could not read the SMS. When forwarding the SMS, users can make use of the contact list stored in their cell phone. However, especially the illiterate participants do not use the contact list on their phones; we oftentimes learned that they had their phone only to make calls to a few numbers they had memorized and mostly received calls from their employer, such as one participant who mentioned: *"I only use my phone so that my boss can call me whenever he needs me. I have no idea how to use any of the functions, and the only number that I can type in and call is the number of my family." (Construction worker, 25)*

This shows that our solution is of limited use for our intended user segment, as there are a large number of illiterate and low-literacy users who would be unable to access voice content via SRLs if they have to comprehend the information from an SMS. Our results showed that accessing the SRLs in this fashion is a solution only for literate users. Unfortunately, there is no way of creating an SRL that uniformly combines a phone number and a LinkCode into one number. Most feature phones offer a way to make a phone call and enter a DTMF number with a certain delay; however, this is not standardized among all phone models. On some phones the letter "p" is used to add a pause and signal the beginning of a DTMF command, while on other phones the "#" sign is used for the same purpose.

To address this issue, we looked beyond feature phones and extended our application to another platform. During our user tests we noticed that even in our user target segment of low-income people, there were a significant number of smartphone owners (four out of 38). We asked these participants informal questions about their usage behavior and their experience using smartphones. While we did not collect enough data to draw quantitative conclusions or report detailed results, there was enough anecdotal evidence for smartphone usage in our user population to make us look more closely at the smartphone situation in India. Therefore, in the next section we present background information about smartphone usage in India and show why it warrants further attention.

7 Background: Smartphones in India

As recent studies of the mobile phone market highlight[7] [8, 9], underprivileged users are slowly switching to smartphones that are becoming affordable for low-income users in developing areas. Contrary to what one might assume, this does not imply that all these people will have Internet access. Despite strong efforts to increase Internet connectivity and campaigns by mobile network providers[8], many Indians are still hesitant to get mobile data contracts. Similar findings have been reported by other researchers looking at developing or emerging countries of the world, such as in Kenya [27] or South Africa [12].

A recent report states that Internet penetration in rural areas went from 2.6% in 2010 to 4.6% in 2012[9], and our encounters with smartphone owners during our study supported these numbers for urban areas as well. Many providers selling smartphones in India offered one month of complimentary Internet access. One participant said that he did not continue using Internet after this month as he saw no benefit: *"I used Internet for a month when I got my smartphone, but there was nothing interesting for me, so I'm not using it anymore." (Security guard, 24)*

The reasons for the lack of interest are many, with one major issue being the language barrier, as most Indians from low-income backgrounds are not literate in English. One of the participants offered additional support for an observation made earlier by Kam et al. [16], saying that the Internet did not offer him enough content in Hindi: *"Most stuff on the Internet that I came across was in English. There is not much you*

---

7 `http://www.ingentaconnect.com/content/routledg/ccon/2012/00000026/00000005/art00003`

8 `http://sharesmartphone.com/2012/03/`

9 `http://www.dnaindia.com/money/report_rural-internet-usage-grows-faster-than-urban_1734825`

can do on the Internet on your smartphone if you can't read English, which is why I stopped using [the Internet]." (Cab driver, 25)

Another participant mentioned that he only used the Internet functionality on his phone to download songs, and that this was too expensive for him since he can get songs from his friends by sharing via Bluetooth for free. This practice is particularly common in communities of underprivileged users in urban areas in India: Bluetooth is used frequently to share content among people, as previous studies show [25]. Similarly, Bluetooth or local Wi-Fi connections are used to share or download applications or other data, accounting for the majority of all application downloads to smartphones in India according to one study [8]. The same study mentions other reasons why users do not purchase data contracts, such as fear of hidden data transfers incurring additional costs, incomplete Internet coverage across different areas especially when leaving popular regions, and unreliable and oftentimes slow connections.

We took the increased proliferation of smartphones into account, developing an Android application that augments the interactive voice application. It relies only on calls and SMS to communicate with the voice-based application and does not make use of data connectivity. Previous research projects, such as Claim Mobile [11], have shown that relying on GPRS can lead to failure for HCI4D applications.

8 Design of Smartphone Application

We chose Android as the platform on which to design the smartphone application. The application uses the same voice-based service that was also used in the feature phone study; it calls the same phone number and receives the same SMS. However, the visual interface augments the phone call and intercepts any incoming SMS based on a regular expression. The final interface with which the user tests were carried out can be seen in Figure 3.

Fig. 3. Screenshots of the Billi Mausi Android application: a) home screen, b) list of recordings, c) single recording selected, d) screen during the phone call

On the home screen (Figure 3 a), the application offers buttons to call the Billi Mausi service or access previously stored recordings. During the phone call, participants see a hint on how to store a recording (Figure 3 b). The application reads the phone's SMS database, populating the list of recordings (Figure 3 c). Instead of the SMS, users see a screen that provides them with a variety of options: listen to the recording, share the recording by entering a phone number, or see whom they shared this recording with already (Figure 3 d). Pressing the "Play" button will initiate a phone call to the voice-based service, but it does not require the user to enter the LinkCode; after a short pause (to make sure that the telephony link has been established), the LinkCode is entered automatically.

Our Android application went through multiple iterations, and preliminary user tests, pilots, and informal focus groups highlighted several issues. In particular, we conducted several pilots with illiterate participants to ensure that the Android application was usable for them even if they had never used a touchscreen phone before. We chose to display as little text as possible but labeled buttons with icons, English text, and Hindi text at the same time. As Hindi symbols are still not officially supported by the Android platform, we added the buttons as images rather than Hindi text in the Android XML files.

For security reasons, telephony access is restricted in the Android SDK: once a call is established, it is decoupled from the application that initiated this call. It is also not generally possible across all devices to access the voice stream, detect DTMF signals, or send DTMF signals during the call. Therefore, we faced many limitations in what the application could do and how to synchronize the call interaction with our Android application. For example, it is currently not possible to implement a "save" button within the application that sends a command similar to *7 to the voice-based application – participants had to manually enter it via the smartphone dial pad. We therefore added a non-modal dialog (an Android *toast*) to display a help message that reminds the user of the *7 command to save a recording (Figure 3 b).

9 Study #2: Sharing with Smartphone

The target audience we aimed for in the second study was the same as in the first study. However, we tried to reach out more for smartphone users from low-income communities to reduce the novelty effect and usability issues due to inexperience. Several issues we observed stemmed from the fact that participants had never used a touchscreen and in particular had no experience with Android interaction – while potential users in the foreseeable future would be smartphone owners and have at least a minimum amount of experience in using it.

Our study comprised 30 participants (different from the first study), who were all male and aged between 19 and 65 (average 32), seven of whom had a smartphone and four of whom used mobile Internet. Typical occupations were similar to the first study, e.g., drivers, shopkeepers, cleaning staff, and security guards. The study protocol was also similar to the one in the first study: we first provided participants with a brief introduction of the study and the system, explained the general purpose, and gave a short demo of the application. We then handed the smartphone over to them and asked them to make a call, record a sentence, and store this recording by pressing *7. After the recording was done and the call ended, participants would be taken to the screen (Figure 3 d). In contrast to the SMS reception in the feature phone study, participants did not need to input anything to access their recording; simply clicking on the "Play" button would call the voice-based application and enter the LinkCode automatically, playing back the recording without any user input required.

At the end of the user test, we asked participants to enter their own cell phone number and forward the recording to their own phone. Similar to the feature phone study, participants read the SMS and were supposed to access their recording. This task was introduced for three reasons: firstly, it enabled us to compare the user set with the one of the first study, making sure that participants had a similar technical background and level of literacy. Secondly, it gave participants the feeling of sharing in a more realistic fashion. Thirdly, it exposed a larger part of the Android application to the participants, ensuring that the whole interaction was possible for our entire target audience. We asked the same questions as for the feature phone study, but concluded the user test by showing them the one-click play button option on the smartphone – if they had not discovered this by themselves in the user test already.

9.1 Results

After the introduction and demo, all participants were able to use the system successfully with a little help depending on their technical background. No participant was completely unable to use the system and no one refused to use it – to some extent, this was due to the interest that the smartphone sparked among participants. Many users were exposed to smartphones for the first time and therefore we did not encounter as much criticism as in the first study.

The SMS task at the end allowed us to roughly categorize their literacy level in the same way as in the first study. The numbers were similar to those in the first study: three participants were illiterate and five low-literate, a combined 27% of all participants in our smartphone study. However, all participants understood the concept and were able to provide us with examples of use cases or ideas for which they would use the sharing functionality.

9.2 Possible Use Cases

Some of the feedback from our participants was similar to that from the feature phone study. In particular, when asked what they could use the service for, many participants came up with ideas similar to the voicemail-like use that was already mentioned in the first study. However, due to the Android interface even illiterate participants were able to use it and had a different experience in this trial, e.g.: *"I'm completely illiterate and can't write an SMS. But with this service I could send my supervisor a message when I'm on sick leave. This service is a very good service for people like us who don't know how to read or write SMS!" (Office support staff, 40)*

While comments like this suggest a one-to-one communication where content is only shared with one other user, some participants envisioned more complex use. One participant said that sharing something with a group of friends could start a conversation, and went further to describe how he would use it for threaded discussion: *"I could start a discussion and my friends could add comments to that and point others to their comments and so on… it could help organize this and provide an overview of the discussion, who replied to whom and so on." (Bus driver, 30)*

The idea expressed in the quote above could be further enhanced by taking advantage of the visual properties of smartphones. For example, by using a smartphone in conjunction with voice-based services, the structure and participants in a discussion could be represented visually, thus further enabling use by and providing value to illiterate users.

10 Discussion

Voice-based telephony applications have always been a promising alternative in technology design for low-income, low-literacy populations that do not have access to the Internet. Over time, these applications have become more powerful and versatile, and designers look into new ways to improve the user experience. We believe that the SRLs on smartphones can provide a new, viable means of interaction for underprivileged people. Based on the insights gained from our interviews and observations during our studies, we highlight opportunities for sharing content and how these can be applied to voice-based systems. As we discuss next, new challenges emerge that need to be addressed; in particular when implementing mechanisms for sharing, but also more general considerations as voice-based services become more popular.

10.1 New Opportunities through Sharing Voice Content

Smartphones for the Illiterate. After we limited the use of displayed text in our application to the bare minimum, participants were able to access SRLs due to its augmentation by the smartphone application. This concept allows for a variety of new interaction mechanics for voice-based systems, as it enables illiterate users to share voice content via SMS. Friscira et al. [14] previously highlighted the value of smartphones for low-literacy users as they enabled them to interact with basic amounts of text. Our study extends this work and enables illiterate, low-income users in developing countries to share any kind of voice content by using smartphones as a vehicle to share SRLs.

"Is that really you?" – Trust in Shared Content. Multiple participants in both studies mentioned that they preferred voice calls over texts as voice cannot be faked easily. Participants felt that voice messages are more authentic. One participant said, looking at the SMS after he listened to the recording: *"It tells me that the sender really is the sender – the combination of voice and [the sender's phone number] is really great!"* By sending an SMS that contains an SRL, the recipient can not only listen to the sender's voice, but also verify his or her number. Smartphones can enhance this effect by giving other visual cues, e.g., personalized icons [14] or symbols representing words [15].

Accessibility of Voice Content. One of the big challenges for voice-based telephony applications is the retrieval of desired content from the huge amount of voice content. Techniques such as searching are not very mature for voice content due to limitation of speech recognition engines for local languages and dialects. Alternate mechanisms of information retrieval such as browsing, different ways of navigation, or directly accessing a specific item in such an application assume more importance [2, 7, 10, 20, 22, 24]. As our second study showed, SRLs offered a quick and simple interaction for accessing voice content, regardless of how deep it is located within the navigation tree.

Multimodal Interfaces. The smartphone version of the Billi Mausi application essentially provides a multimodal interface to the users (visual and voice). None of the participants complained about having to deal with multiple modes; rather the multimodality was appreciated. This acceptance opens up opportunities for designing smartphone applications for this user segment that can rely on multiple modes of interaction simultaneously.

10.2 Challenges When Designing Applications for Sharing

Financial Concerns. Among underprivileged people, fear of hidden costs is a major concern when using telephony services [8]. In our application, a short popup message informed the user about SMS charges after they hit the "Share" button, keeping the costs transparent. If an application not only establishes trust in its affordability, but actually *saves* the user money, it may evolve into a thriving service, as participants pointed out, e.g.: "*If this service was free, I would contact a lot of people I usually don't have contact with.*" A service does not need to be free in order to be used – if SRLs would be used to facilitate quick access to valuable content such as a health advisory or farming advice, offering a way to save money, this challenge can turn into an advantage. Especially among low-income users in developing countries, financial feasibility can be the difference between success and failure of an application.

Interface Customization for Voice-Based Services. Participants sometimes mentioned a desire for customization in several stages of the interaction. Even though these comments were a response to SRLs, they apply to voice-based services in general. One participant mentioned that he would like to change the store command *7 to his *"lucky number"*. This brings up the question how to support customization and maintaining necessary consistency. The platform that our application was built upon used the * key to expose universal commands and 7 was allocated for saving a link to a content resource while other keys had different assignments. This follows established design guidelines for IVRs [6].

Recipient Unknown – Design for Diverse Technology. As smartphones become increasingly popular among underprivileged users, designers of voice-based systems have many options for interaction design, such as using visually augmented SMS or automatically dialed SRLs. However, these opportunities come with a challenge: the recipient's phone is an unknown device. It may be a feature phone, it may be a smartphone, or it may be a smartphone that is not running the same application as the sender. When designing voice applications, different client usage scenarios need to be accounted for. For instance, in our study, the same voice application worked for users with a feature phone as well as through a multimodal interface via a smartphone.

11 Conclusion

We extend research on interactive voice-based telephony systems by proposing SRLs, a new means of sharing voice content. Our studies show that underprivileged users are able to use the service and understand the sharing functionality. At the same time, they exposed new issues that need to be taken into consideration when developing for sharing of voice content. Furthermore, we only investigated the sharing of user-generated content and looked at first-time users – but our findings suggest opportunities for a variety of different services and further investigations as we venture into a previously unexplored domain. Literacy issues pose one of the most difficult challenges for designing for underprivileged populations, and we provide evidence that smartphones can help to address these issues and even enable illiterate users to take advantage of the benefits SRLs offer, increasing the possibilities of sharing among underprivileged users. Looking back at our scenario in the introduction of this paper, Raj could use the functionality implemented in the Billi Mausi application to request information from a physician, and once he receives the answer, he could forward the information to his cousin by simply forwarding the SMS containing the SRL.

We believe that sharing of voice content can greatly enhance information dissemination among the underprivileged and that SRLs enable new ways of interacting with voice-based services. Therefore, our next steps are to apply SRLs to a variety of services that are more content-driven and reflective of current voice-based information systems in use today than Billi Mausi, which was built for the purpose of studying the basic sharing mechanics. We intend to design applications that lend themselves to long-term deployment studies in order to investigate the full potential of SRLs.

Acknowledgements. We would like to thank Elaine M. Huang and Gunnar Harboe for their invaluable comments and input on early drafts of this work. We also thank Meghna Singh, Jyoti Grover, and Monia Puri for helping implement the Billi Mausi application.

References

1. Agarwal, S., Kumar, A., Nanavati, A.A., Rajput, N.: Content creation and dissemination by-and-for users in rural areas. In: Proc. ICTD 2009 (2009)
2. Ajmera, J., Joshi, A., Mukherjee, S., Rajput, N., Sahay, S., Shrivastava, M., Shrivastava, K.: Two Stream Indexing for Spoken Web Search. In: Proc. WWW 2011 (2011)
3. Anokwa, Y., Smyth, T., Ramachandran, D., Sherwani, J., Schwartzman, Y., Luk, R., Ho, M., Moraveji, N., DeRenzi, B.: Stories from the Field: Reflections on HCI4D Experiences. Information Technologies & International Development 5(4) (2012)

4. Awaaz De Infosystems Pvt. Ltd. Awaaz.De (2012), http://www.awaaz.de/
5. Beaton, J., Kumar, R.: Indian Cultural Effects on User Research Methodologies. In: Ext. Abstr. CHI 2010 (2010)
6. Blanchard, H.E., Lewis, S.H.: Voice messaging user interface. In: Gardner-Bonneau, D. (ed.) Human Factors and Voice Interactive Systems. Kluwer (1999)
7. Borodin, Y., Dausch, G., Ramakrishnan, I.V.: TeleWeb: Accessible service for Web Browsing via Phone. In: Proc. W4A 2009 (2009)
8. ConsumerLab, Ericsson. Emerging App Culture. An Ericsson Consumer Insight Summary Report (2012)
9. Convergence Catalyst. India Smartphone Outlook 2012 (2012)
10. Danesha, K.A., Rajput, N., Srivastava, K.: User Driven Audio Content Navigation for Spoken Web. In: Ext. Abstr. MM 2010 (2010)
11. Densmore, M.: Claim Mobile: When to Fail a Technology. In: Proc. CHI 2012 (2012)
12. Dhir, A., Moukadem, I., Jere, N., Kaur, P., Kujala, S., Ylä-Jääski, A.: Ethnographic Examination for Studying Information Sharing Practices in Rural South Africa. In: Proc. IARIA 2010 (2012)
13. Esoko, Esoko – For farmer (2012), http://www.esoko.com/about/index.php
14. Friscira, E., Knoche, H., Huang, J.: Getting in touch with text: designing a mobile phone application for illiterate users to harness SMS. In: Proc. DEV 2012 (2012)
15. Heimerl, L., Honicky, R.J., Brewer, E., Tapan, P.: Message phone: A user study and analysis of asynchronous messaging in rural uganda. In: Proc. NSDR 2009 (2009)
16. Kam, M., Ramachandran, D., Devanathan, V., Tewari, A., Canny, J.: Localized Iterative Design for Language Learning in Underdeveloped Regions: The PACE Framework. In: Proc. CHI 2007 (2007)
17. Kumar, A., Tewari, A., Shroff, G., Chittamuru., D., Kam., M., Canny, J.: An Exploratory Study of Unsupervised Mobile Learning in Rural India. In: Proc. CHI 2010 (2010)
18. National Statistical Commission, Government of India. Report of the Committee on Unorganised Sector Statistics (2012)
19. Patel, N., Chittamuru, D., Jain, A., Dave, P., Parikh, T.S.: Avaaj Otalo - A Field Study of an Interactive Voice Forum for Small Farmers in Rural India. In: Proc. CHI 2010 (2010)
20. Ramakrishnan, I.V., Stent, A., Yang, G.: Hearsay: enabling audio browsing in hypertext content. In: Proc. WWW 2004 (2004)
21. Raza, A.A., Haq, F., Tariq, Z., Pervaiz, M., Razaq, S., Saif, U., Rosenfeld, R.: Job Opportunities through Entertainment: Virally Spread Speech-Based Services for Low-Literate Users. In: Proc. CHI 2013 (2013)
22. Reddy, H., Annamalai, N., Gupta, G.: Listener-controlled dynamic navigation of voiceXML documents. In: Miesenberger, K., Klaus, J., Zagler, W.L., Burger, D. (eds.) ICCHP 2004. LNCS, vol. 3118, pp. 347–354. Springer, Heidelberg (2004)
23. Sambasivan, N., Weber, J.S., Cutrell, E.: Designing a Phone Broadcasting System for Urban Sex Workers in India. In: Proc. CHI 2011 (2011)
24. Srivastava, S., Rajput, N., Mahajan, G.: SWAicons: Spoken Web Audio icons: Design, Implications and Evaluation. In: Proc. CSCW 2012 (2012)
25. Smyth, T.N., Kumar, S., Medhi, I., Toyama, K.: Where There's a Will There's a Way: Mobile Media Sharing in Urban India. In: Proc. CHI 2010 (2010)
26. White, J., Duggirala, M., Srivastava, S., Kummamuru, K.: Designing a Voice-based Employment Exchange for Rural India. In: Proc. ICTD 2012 (2012)
27. Wyche, S.P., Murphy, L.L.: Dead China-Make. Phones Off the Grid: Investigating and Designing for Mobile Phone Use in Rural Africa. In: Proc. DIS 2012 (2012)

Adding Vibrotactile Feedback to Large Interactive Surfaces

Julian Seifert[1], Markus Packeiser[2], and Enrico Rukzio[1]

[1] Ulm University, Ulm, Germany
{julian.seifert,enrico.rukzio}@uni-ulm.de
[2] University of Duisburg-Essen, Essen, Germany
markus.packeiser@uni-due.de

Abstract. Interactive surfaces and multi-touch tables are increasingly available outside academic contexts, and are entering, for instance, work or educational contexts. A large variety of applications exists for a multitude of tasks. For interacting with these applications, existing interaction concepts are often directly mapped to the multi-touch surface, which is often limited by physical constraints. For instance, to enter text on an interactive surface, most often a virtual keyboard is used. However, users cannot feel when, for instance, they have accidentally pressed two keys at the same time. Research on mobile devices has identified vibrotactile feedback as an effective means to support users when interacting with touch screens. In this work, we present results of an experiment in which we investigated whether typical tasks (e.g., typing text, drag-and-drop of items) on interactive multi-touch surfaces can be supported by providing vibrotactile feedback directly on the surface[1]. We compared direct feedback with distal feedback provided on the user's body, as well as their combination. Surprisingly, our results show that all compared variants of vibrotactile feedback had no significant positive effect on the task performance. Yet participants rated tactile feedback significantly higher regarding interaction support and subjective speed compared to no provided feedback.

Keywords: Interactive surfaces, vibrotactile feedback, touch-based interaction, multi touch.

1 Introduction

Recently, large interactive surfaces have begun to come out of the labs and are entering diverse usage contexts such as domestic, or workplace environments. Reasons are maturing technologies and progress in interaction research, which enable an increasing number of use cases wherein direct multi-touch on interactive surfaces is the main interaction style. Touch-based interaction is commonly considered to be natural [6] as users can interact with virtual objects as they expect it based on their experiences with real physical objects [7]. However, depending on the task, the absence of

[1] A demo video is available at http://youtu.be/0DzTtBTeglQ

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 507–514, 2013.

corresponding tactile feedback of virtual objects decreases user experience and performance (which is in particular a challenge in the field of virtual reality [2]).

This challenge of missing corresponding tactile feedback applies also to touch-based interaction with hand-held mobile devices such as smartphones. One solution that has been identified to significantly increase the typing performance on virtual keyboards is vibrotactile feedback [3]. That is, users benefit from short vibrotactile sensations when touching a button, indicating that an effect has resulted from the input.

Evidently, vibrotactile feedback is a promising candidate for increasing performance and accuracy of touch-based interactions on large interactive multi- touch surfaces. However, it is not obvious how differences to hand-held devices (e.g., multi-touch interactions, screen size, task-type) influence the positive effect of vibrotactile feedback.

In this work, we contribute findings of a user study investigating the effect of vibrotactile feedback during interaction with interactive surfaces enabling multi- touch. The study compared direct vibrotactile feedback on the surface with distal feedback on the user's wrist [11], as well as the combination of both. As a base- line, no vibrotactile feedback was given. Surprisingly, our findings indicate that vibrotactile feedback provided directly on a surface does not increase the users' performance in a typing task as observed on hand-held devices. In fact, none of the tested feedback conditions results in a significant improvement compared to conditions wherein no feedback was provided. However, participants expressed that vibrotactile feedback (in any form) is helpful and preferred over no feedback.

In the remainder of this paper, we discuss related works, introduce our experimental setup, and discuss the results and observations gained during the experiment.

2 Related Work

Early work on vibrotactile feedback was done in the field of virtual reality (VR). In this context, the absence of tactile feedback while interacting with virtual objects was identified as a challenge and different approaches for providing force feedback (e.g., [4]) and tactile feedback [5] have been introduced.

For interaction with interactive surfaces such as touch screens, the challenge is similar: users can touch objects displayed on the screen. However, it is always the same sensory experience.

Similar as in VR, additional input devices can be used which provide vibrotactile feedback when necessary. For instance, the *HapticPen* enables one or multiple users to interact with a touch screen and experience individual feed- back [8]. Marquardt et al. introduced the *Haptic Tabletop Puck*, a mouse-like device providing feedback which the user moves on the tabletop to experience different surface qualities [10]. McAdam and Brewster investigated the effect of distal feedback on typing tasks using a small touch screen [11]. Conceptually, distal feedback is provided by the users' personal mobile phone and its built-in vibration alarm. In their study, they used external vibration actuators placed on the user's body. Their study results indicate that distal feedback improves users' typing performance compared to when no feedback was provided.

Among the approaches that include an external device which provides the vibrotactile feedback, all have in common that the interaction with objects is less or not at all natural because direct touching is not possible. Hence, approaches that provide tactile feedback directly on the surface are capable of maintaining the natural interaction. Brewster et al. investigated the effect of vibrotactile feedback provided directly on the touch screen of mobile phones [3]. They found that such feedback has a significant positive effect on users' performance.

Another approach for enriching the tactile experience on interactive surfaces was presented by Bau et al. with *TeslaTouch* [1]. TeslaTouch is based on electro- vibration and hence, no physical movement of the surface is required to create the feedback. However, the electrovibration changes the force of attraction between the touching finger and the surface, which results in a specific perceived friction when moving the finger. Therefore, the feedback can only be provided for dragging interactions.

In this work, we examine the approach of providing tactile feedback through attaching vibration actuators at the bottom side of a large interactive surface which allows us to consider not only the dragging of objects but also tasks such as typing, in which the fingers of the user only touch the surface to trigger a button without dragging. Further, the size of the interactive surface is of relevance when considering multi-touch interactions.

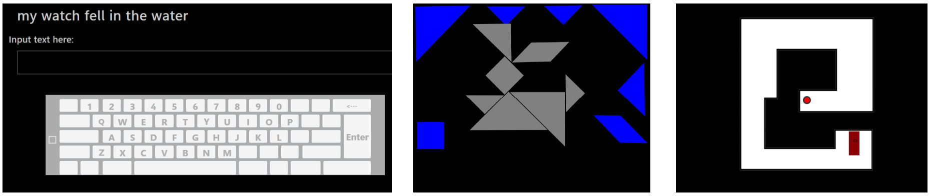
3 User Study

We designed and conducted an experiment1 in order to investigate how vibrotactile feedback effects interaction with an interactive surface in comparison to settings with distal feedback and without any feedback. In the following, we introduce the design, tasks, and apparatus of the user study.

Design. We applied a within-subject design. The independent variable was the feedback: direct feedback on the interactive surface, distal feedback provided on the user's body, and combined direct and distal feedback. Finally, as baseline, no feedback was provided.

In counterbalanced order, participants used each of the four feedback variants to complete a series of tasks, which were also presented in counterbalanced order. While interacting with the interactive surface, participants were standing and allowed to move around the table. After each round with a feedback variant, participants filled in a questionnaire rating the previously tested feedback variant.

Tasks. Similar to previous work, we argue that typing is not the primary and only activity that is likely to be performed with an interactive surface. Hence, we included two additional tasks to test the effect of each feedback variant. These required drag-and-drop actions, the use of multiple fingers at the same time, and a steering task.



(a) (b) (c)

Fig. 1. Graphical user interfaces of the task applications: (a) a typing task, (b) a puzzle, and (c) a steering task

1. *Typing task.* In the typing task, participants were asked to type six English sentences during each feedback condition. The sentences were taken from a standard set designed for typing performance testing [9]. The sentences are of equal length and complexity and do not contain special characters or punctuation marks. The user interface of the typing task application shows three items: the given sentence to type, a text field, and a virtual QWERTY keyboard (see Fig. 1a). Users can move the virtual keyboard to a preferred position in which they find it comfortable to type. Vibrotactile feedback (if available) was provided when hitting a character key on the virtual keyboard.

   The collected data consisted of the task completion time, the number of corrections, and the number of errors.

2. *Puzzle.* In the puzzle task, participants were given a number of shapes on the interactive surface which had to be matched with predefined positions which were indicated by gray placeholders (see Fig. 1b). Vibrotactile feedback (if available) was provided when the user managed to place an item on its target (considering location and rotation within a specific tolerance). The puzzle was solved when all items were positioned correctly.

   The collected data were the task completion time and the number of times an item was released outside the accepted tolerance.

3. *Steering task.* In the steering task, participants were asked to steer a rectangle through a given path without touching the borders of the path (see Fig. 1c). When they reached the end of the path, participants had to follow the path back to the starting point. Vibrotactile feedback (if available) was provided when the distance between the path border and any corner point of the rectangle was less than a specific threshold.

   The collected data encompassed the task completion time and the number of times the rectangle shape accidentally hit the border of the path.

Apparatus. The apparatus system for the experiment consists of an interactive surface which is augmented with four vibration actuators (see Fig. 2a). The interactive surface is based on frustrated total internal reflection (FTIR) for sensing multiple touch events simultaneously. The surface display is based on a rear projection (1280×1024 px), controlled by a computer which runs surface applications.

To generate the vibrotactile feedback that can be perceived when touching the surface of the multi-touch table, four 50 Watt low-frequency speakers (20-80 Hz) are mounted inside the interactive surface. They are powered by two 100 Watt amplifiers that are connected to the surface computer. In order to provide vibrotactile feedback, the running surface application plays an audio file which specifies the vibration pattern (a square wave was empirically found to produce a strong tactile sensation on the surface). Similar to the direct feedback, the distal feedback is generated using a vibration actuator which was attached to the participant's wrist (see Fig. 2c).

(a) (b) (c)

Fig. 2. The apparatus systems: (a) An interactive surface using rear projection. (b) Four 'bass pumps' mounted inside the interactive surface in order to generate low- frequency vibrations. (c) C2 vibration actuator to provide distal feedback.

Participants. Through an email list, we recruited 18 participants (10 males and 8 females) aged between 21 and 31 years. Participants were randomly selected. All but one participant (a teacher) were students with diverse backgrounds (technical and humanities). Participants received 10 Euro after the study as a reward.

Hypotheses. Overall, we hypothesized that vibrotactile feedback has a positive effect on the interaction. In particular:
H0: Direct vibrotactile feedback provided by the surface leads to better (faster task completion and less errors) performance during interaction (typing, puzzle, and steering) compared to no feedback and to distal feedback.
H1: Combined distal and direct feedback outperforms their separate effect (faster interaction and less errors).
H2: Direct feedback is subjectively preferred over distal and combined feedback by users.

4 Results

We hypothesized that vibrotactile feedback has a positive effect on the task completion time (*H0*). To our surprise, the mean task completion times for the three tasks are very close to each other. As figure 3 shows, in the typing task, the average completion time for the condition 'no feedback' is higher (M=72.7s) com- pared to distal (M=68.4s), direct (M=66.6s), and combined feedback (M=64.8s). Yet testing with a one-way repeated-measures ANOVA shows that differences are not significantly different (F(3; 51)=1.76; p=.17). Moreover, for the puzzle and the steering task, the

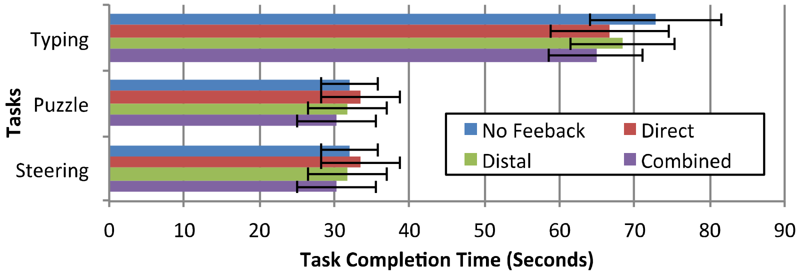


Fig. 3. Task completion times for the tasks 'typing', 'puzzle', and 'steering' (Error bars show the standard deviation)

direct feedback resulted in the longest completion times. This raises the question why direct vibrotactile feedback has a significant positive effect on typing performance when using a mobile device (see [3]), however, not when applied on a large surface.

Further, we compared the number of fixes (how many times the backspace key was pressed). Overall, there is a significant effect (F(1.77; 29.99)=6.77; p=.005): Direct feedback (M=4.22, SD=4.70), distal feedback (M=4.28, SD=3.14), and combined feedback (M=4.72, SD=4.01) equally often required the user to correct typing errors, followed by no feedback (M=10.05, SD=8.92). A pairwise comparison shows that only direct feedback requires significantly less fixes com- pared to no feedback (p=.04).

The minimum string distance (i.e., Levenshtein distance) which gives an estimation of the number of errors made, did not show significant differences between the four feedback types (F(3; 51)=.38; p>.05).

During the task of solving the puzzle, participants performed numerous drag- and-drop actions. This type of interaction differs from typing as the user's fingers are in contact with the surface for longer times. Hence, vibrotactile feedback should help to place the puzzle pieces in place correctly. However, looking at the average number of falsely placed pieces, it shows that combined feedback resulted in the highest number of errors (M=5.9; SD=3.3), followed by distal (M=4.8; SD=3.7), no feedback (M=4.3; SD=1.7), and direct feedback (M=4.1; SD=1.7), yet the differences are not significant (F(3; 51)=.81; p>.05).

Comparing the amount of errors made during the steering task shows no significant differences (F(3, 51)=2.46; p>.05). However, the numbers show a tendency indicating that feedback is helping: no feedback (M=4.89, SD=4.32) compared to direct (M=2.72, SD=2.73), distal (M=2.61, SD=2.35), and combined feedback (M=2.83, SD=3.22).

These results contradict to our hypothesis H0, which consequently has to be rejected. Further, the combination of direct and distal feedback does not result in better results. Consequently, H1 has to be rejected as well.

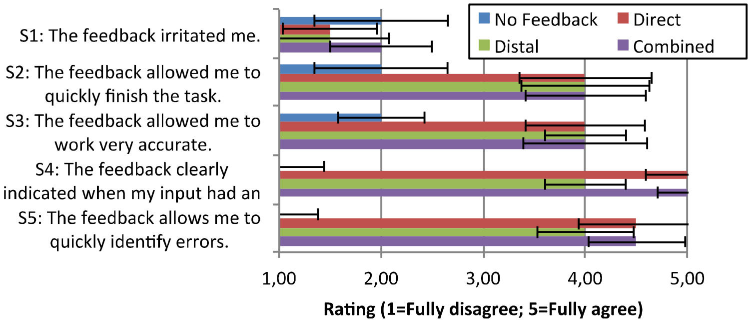


Fig. 4. Subjective ratings of the four feedback conditions (Bars show medians; error bars show standard deviation)

After each task, participants rated the feedback type via different statements (see Fig. 4) on a 5-point scale (1=fully disagree). Regarding the statement S1 ('The feedback irritated me.'), participants did not rate the feedback types significantly different ($\chi^2(18)=5.8$; $p>.05$). For all other statements, significant differences were reported, as illustrated in Fig. 4. However, interestingly the only significant differences are those between conditions in which feedback was given and the baseline condition in which no feedback was provided, as shown by pairwise comparison. For instance, in statement S3 ('The feedback allowed me to quickly finish the task.'), all three conditions in which vibrotactile feedback was provided were rated the same on average (Mdn=4.0). In statements S4 and S5, distal feedback was rated lower compared to direct and combined feedback. As these differences are not significant, it might at least indicate that the presence of direct feedback is appreciated by users. These subjective ratings indicate that from the users' point of view, it only matters that feedback is provided at all – not how this is implemented. Accordingly, hypothesis H2 cannot be confirmed.

5 Discussion and Conclusion

In this work, we investigated the effect of vibrotactile feedback on interaction with large interactive surfaces. Feedback provided directly on the surface, distal, or combined, did not have an effect on the users' performance. In fact, this result is surprising as previous work showed that vibrotactile feedback significantly improves speed and accuracy.

One obvious explanation for the small objectively measured effect is that performing touch and drag-and-drop interactions on an interactive surface already causes strong tactile sensations. Hence, a vibrotactile feedback has to be even stronger compared to this resulting sensation at the users' fingertips. In the case of mobile handheld devices, we hypothesize that the hand holding the device is mainly responsible for feeling the vibrotactile feedback, rather than the finger used for input.

In addition, the results of this experiment depend strongly on the used FTIR technology, which requires users to perform stronger touches compared to a capacitive touch screen. Hence, future work should investigate if and how vibrotactile feedback is beneficial on capacitive surfaces such as the Samsung SUR40. The data measured during the experiment, such as task completion time different error values, did not support our initial hypothesis that direct feedback has a positive effect on the performance. However, users reported that any given type of vibrotactile feedback is valuable to them as they felt that they performed tasks faster.

As the tested feedback types (direct, distal, and combined) do not cause considerable differences in terms of speed (task completion) and accuracy (errors), the decision regarding which type of feedback to offer depends mainly on the given application design. For instance, in multi-user games it might be interesting to provide a *common* feedback for all players directly on the surface. At the same time, private feedback could be provided via distal means.

Acknowledgement. This work has been conducted in the Emmy Noether research group Mobile Interaction with Pervasive User Interfaces funded by the German Research Foundation.

References

1. Bau, O., Poupyrev, I., Israr, A., Harrison, C.: Teslatouch: electrovibration for touch surfaces. In: UIST 2010, pp. 283–292. ACM (2010)
2. Benali-khoudja, M., Hafez, M., Marc Alex, J., Kheddar, A.: Tactile interfaces: a state-of-the-art survey. In: Int. Symposium on Robotics, pp. 721–726 (2004)
3. Brewster, S., Chohan, F., Brown, L.: Tactile feedback for mobile interactions. In: CHI 2007, pp. 159–162. ACM (2007)
4. Brooks Jr., F.P., Ouh-Young, M., Batter, J.J., Jerome Kilpatrick, P.: Project GROPEHaptic displays for scientific visualization. In: SIGGRAPH 1990, pp. 177–185. ACM (1990)
5. CyberGlove Systems LLC. Overview (2013), `http://www.cyberglovesystems.com/products/cybertouch/overview` (last access: March 15, 2013)
6. Dietz, P., Leigh, D.: Diamondtouch: a multi-user touch technology. In: UIST 2001, pp. 219–226. ACM (2001)
7. Ishii, H., Ullmer, B.: Tangible bits: towards seamless interfaces between people, bits and atoms. In: CHI 1997, pp. 234–241. ACM (1997)
8. Lee, J.C., Dietz, P.H., Leigh, D., Yerazunis, W.S., Hudson, S.E.: Haptic pen: a tactile feedback stylus for touch screens. In: UIST 2004, pp. 291–294. ACM (2004)
9. MacKenzie, I.S., Soukoreff, R.W.: Phrase sets for evaluating text entry techniques. In: CHI EA 2003, pp. 754–755. ACM (2003)
10. Marquardt, N., Nacenta, M.A., Young, J.E., Carpendale, S., Greenberg, S., Sharlin, E.: The haptic tabletop puck: tactile feedback for interactive tabletops. In: ITS 2009, pp. 85–92. ACM (2009)
11. McAdam, C., Brewster, S.: Distal tactile feedback for text entry on tabletop computers. In: BCS-HCI 2009, pp. 504–511. British Computer Society, Swinton (2009)

Analysis and Visualization of Interactions with Mobile Web Applications

Paolo Burzacca and Fabio Paternò

CNR-ISTI, HIIS Laboratory
Via Moruzzi 1, 56124 Pisa, Italy
{paolo.burzacca,fabio.paterno}@isti.cnr.it

Abstract. We present a novel solution for intelligent analysis and visualization of user interactions with Web applications through mobile devices in order to help identify usability issues. The proposed tool is also able to support comparison of optimal use with actual user interactions. We also report on an example application of our tool to the evaluation of a real mobile Web site.

Keywords: Tools for Remote Evaluation, Web Applications, Mobile HCI.

1 Introduction

Logging tools for supporting usability evaluation have long been investigated in the context of desktop access. However, such tools cannot be simply re-proposed as they are for mobile devices, given their specific characteristics in terms of relevant events and usability issues. Indeed, mobile devices are becoming ever richer in terms of sensors, such as accelerometers, GPS, and so on. In addition, because of the limited screen size mobile devices have specific usability problems, such as tedious activities in zooming in and out for viewing the desired piece of information or touch/based interactions that select the wrong elements [3].

Remote evaluation supports the analysis of the user behavior in real contexts of use, and this is important with mobile applications that can be accessed in varying contexts. Various tools for capturing logs of user interactions have been proposed but they were limited in various aspects. Thus, there is a need for improvements in tools able to support remote usability evaluation, in particular when mobile devices are considered. For example, WebQuilt [6] only performed proxy-based logging and, thus, it was not able to gather detailed information regarding user interactions and the surrounding environment. Other approaches exploiting client-based logging [1][5] provided reports of the data collected not easy to interpret. The evaluation of the interactions with mobile application by comparing task models with the logs associated with real use was investigated in [5]. In this paper we propose a novel approach in which the logs of the actual use are compared with an optimal log created by a user interface expert.

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 515–522, 2013.

2 The Environment for Remote Evaluation

The proposed environment aims to support remote evaluation of Web applications accessed through any type of device. In this paper we focus on the specific aspects related to mobile access. In order to perform a usability evaluation there are three main phases: *preparation*, during which the evaluators define the tasks that should be performed during the test and provide examples of optimal performance of such tasks; *test*, during which the users access the application from any device through a proxy server that includes the logging scripts and stores the logs in the backend; and the final *usability analysis*, which is based on the information collected beforehand and exploits visualizations that facilitate the identification of potential usability problems.

The usability server includes a proxy through which users should access the Web applications. Such logging can detect any type of event, even events defined by the evaluators. The backend part supports functionalities able to provide useful visualizations of logs through interactive, configurable timelines, and intelligent comparison of actual behaviour with optimal interaction sequences.

One of the activities to carry out during the preparation phase of a user test by the evaluators is the creation of the task list. For each task the evaluator has to provide the name, description, if it has to be started from a specific page then its URL should be specified, whether it can be skipped, whether its performance depends on some other task. For the interaction with end users the choice was to interfere as less as possible with the access to the applications. Thus, the only interaction is given by the task panel, which provides some basic instructions at the beginning of the session, and is used to indicate the start and the end of a task performance, and the next task to accomplish. Thus, the user starts the test through the public user interface, and the proxy includes the control panel, which contains instructions regarding the task, and it allows users to start/finish/jump the task, with minimal intrusiveness.

In our environment the logging is performed through an infrastructure that has been designed for this purpose. We use a log model in which each event is defined by <timestamp, element, id, event, which, extra>, *timestamp* indicates when the event occurred; *element* is the DOM element associated to the event; *id* is the unique identifier of the element; *event* indicates the event type (e.g. click, touchstart, scroll,...); *which* indicates the value of the event object (e.g. what character has been entered when pressing a key or what mouse button has been pressed); *extra* is used to provide additional information, when necessary, such as URL page and window size for the on load page event, GPS coordinates for GPS events, screen coordinates for mousemove or click events. The events recorded include: all the standard events, touch events, events related to accelerometer and GPS, DOM Mutation events, basic semantic events (pageview, starttask, closetask, …). We also manage semantic events that are used in our environment in order to provide more information that can be useful for the evaluators. Thus, we have a set of basic semantic events mainly associated with the task panel access or the task performance. In addition, it is possible to define custom semantic events, which are composition of basic events or standard events on specific parameters with specific event names.

3 The Timelines and Other Visual Information

Timelines are used to represent information gathered from the logs. Previous solutions [1] have provided rather static and limited representations of such timelines, we have thus designed new solutions to represent and manipulate them.

The tool allows evaluators to access the list of available timelines from a given user test ordered by time of the corresponding log files. The first timeline is that generated by the evaluator to store an optimal performance of the given task (Figure 1 shows an example), and it is used as a comparison reference for assessing the performance of the various users, and identifying possible usability problems. Each timeline expands temporally from left to right. For each timeline some useful data are shown: an identifier, an incremental number associated with the corresponding task, when the corresponding log was detected, and the overall duration of the user session. In addition, some information regarding the device used is provided. For this purpose, during the user test the environment takes the user agent and uses it to retrieve more detailed information from a Device Description Repository (we have used WURFL[1]).

Our environment provides the evaluator with various interactive functionalities useful to ease the comparison of logs and their analysis. The timeline visualization is completely dynamic and interactive. All timelines and events can be selected in order to apply specific operations to them. The initial representation of the timeline shows all the events in the log. This can be confusing and complicate the analysis because not all the events may be relevant. Thus, the tool provides evaluators with the possibility of interactively filtering the events through a panel with one check-box for each event category (e.g. form events, keyboard events, touch events, ..).

The events shown are represented through a label and a coloured small circle. The timelines also include some vertical black lines among the events indicating when a page change has occurred. When an event is selected then the corresponding information (e.g. when it occurred, in which element it occurred, ..) is shown. It is also possible to search for a given event, in this case only the occurrences of such event are shown. In case multiple timelines are shown then they are lined according to the first occurrence of the searched event. In order to facilitate the comparison of various logs the tool allows the evaluator to line up the timelines according to when users started to navigate on a given page as well (Figure 1 shows an example). This facilitates the comparison of how different users interacted within the same page. The tool has a zoom feature for adjusting the time scale, thereby allowing the evaluator to expand or contract the events in the timelines. Indeed, it often happens to find many events grouped in very small time intervals, and thus they have overlapping, confusing representations, which can be improved through this feature.

While the timelines provide an effective interactive dynamic representations of the users' sessions, we also introduced an event analyzer with the goal to provide an overall summarizing representation of what happened during the test. The evaluator can select an event category and the tool shows a list of bar charts showing the number of occurrences of such events for each session.

---

[1] `http://wurfl.sourceforge.net/`

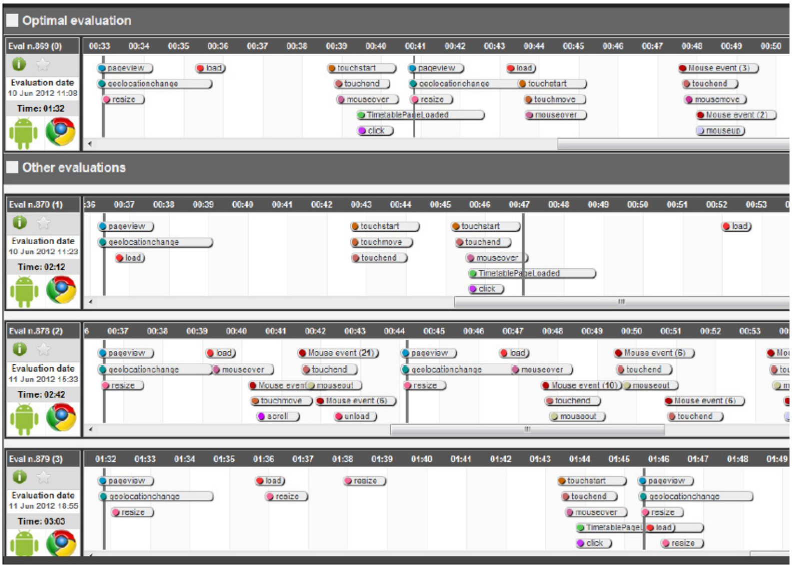


Fig. 1. An example of a set of timelines aligned according to one page access

The tool also provides the possibility of more explicit comparisons among various sessions. For example, it is possible to compare through bar charts the navigation time across the various pages that have been accessed by the considered sessions.

In addition to time-related information, another useful piece of information is the path followed by the user during the navigation. It is not immediate and intuitive to extract such information from the timelines. Thus, our environment also provides the possibility to show visual storyboards, which also easily highlight possible cycles during the navigation. As you can see in Figure 2, each page is represented by a rectangle with an identifier, the visit time, and an arrow indicating the next page accessed.

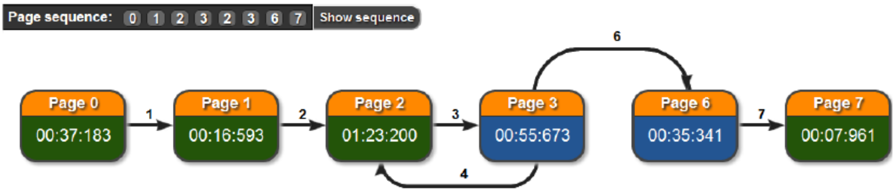


Fig. 2. An example representation provided by the visual storyboard

If a page is visited multiple times then the time is the sum of all the visit times. The arrows are annotated with numbers indicating their temporal order. The rectangles indicating the pages navigated have different colours depending whether they have been accessed in all the considered sessions or not. For each page that can be analysed it is also possible to show its screen dump captured during the user session, graphically annotated in order to indicate where the user interactions occurred.

4 Automatic Log Analysis

In our environment we have also investigated the possibility of providing more automatic support in the log analysis, still through comparison of the user logs with the optimal log initially provided by the evaluator. For this purpose, we have applied the Sequence Alignment Method (SAM) to the comparison of logs of user interactions. Previously, it has been applied to identify navigation patterns in Web applications [2]. We have investigated how to exploit it in order to help identifying potential usability problems. In general, the SAM approach aims to calculate the differences among two sequences of elements. The difference between sequences is represented by the number of operations necessary to make them equal, where each operation has a different weight, which depends on its importance. The operations available for equalizing two sequences are reordering, insertion, and deletion.

In our logs we have information related to both pages accessed and events generated. Thus, when we apply the SAM analysis we distinguish the distance in terms of both types of elements. Actually, in the events we also distinguish between custom semantic events and the others. In the end, the distance between two logs is calculated by summing the distance of these three sequences, where the composing distances have different weights to reflect their different importance for the usability analysis. Indeed, the differences in pages accessed is more important, since if the users have visited pages different from those accessed in the optimal session, then they have probably not accomplished their task correctly. For the custom semantic events, most analysis is limited to check whether they occur in the sequence considered. For the other events they are considered the same if they occur in the same page and on the same user interface element. The weights associated with the elements compared and the type of SAM operation play an important role in determining the final result. Since it is difficult, and probably not possible, to find general weights that always provide the most meaningful results, we decided to allow evaluators to customize them. For example, in the page sequences comparison an evaluator should consider the application type considered and its navigation model: if the task can be accomplished through different paths within the Web application then the weights for the page sequences differences should be low because some differences are probably not indicating any particular problem. The tool also provides some useful information regarding the differences among the sequences. For example, in the case of the page sequences the tool also indicates the pages that differ in the two sessions. As Figure 3 shows, for such pages it also indicates the page previously accessed, which is useful to allow the evaluator to understand from which page the user has taken the wrong path.

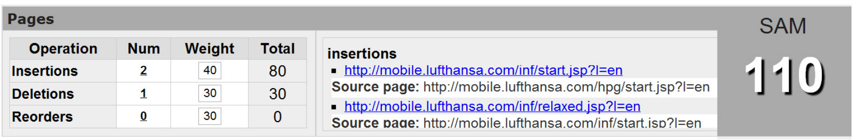


Fig. 3. An example of SAM analysis report on the pages accessed

5 An Example Application

In order to validate our environment we considered an example application, the mobile version of an airline Web site[2]. We first asked some end users to test it through our logging infrastructure and report on its usability issues, then we asked some evaluators to use the information provided by our tool to identify the application usability problems in order to check whether they would have found similar problems.

The application was accessed remotely by the users without any direct support from the evaluators. Users only received at the beginning some information by email about the purpose of the test and how to start it. Users were free to choose the mobile device used for the test and the exact time and location to carry it out.

The test was carried out by seven users, the average age was 32, and they had not used our tool beforehand. The test consisted in performing five tasks chosen in such a way to consider the various aspects characterising Web access. For each task we defined some custom semantic events to have more control in the analysis of the user logs. The first task was to search for a flight from Paris to London with intermediate stop in Frankfurt for a precise day in economy basic fare. In this way the users had to choose among various options and go through various steps to complete the task. We defined a custom semantic event with the selection of the One Way option to better check that users selected it. The second task was to find information about the lounge in Frankfurt. Here the most difficult part was to find the page regarding lounges. Actually, for this purpose two different paths across the Web site were possible, and we included two custom semantic events to easily identify which one was taken. The third task was to find the time for a flight from Munich to Cracow in a given date. The flight should have been direct, and the last available on the indicated date. Also in this case we introduced a couple of custom semantic events. The fourth task was to find how to reach the Frankfurt airport from Strasburg by bus. In this task there was no form to fill in. The fifth task was to find information about the Boeing 737-300. We introduced one custom semantic event to check the path followed.

Regarding the first task, some users complained that the Web site had the one way option clearly visible and highlighted, thus erroneously leading them to think that it was selected, and they only realized that this was not the case when the application asked them to select the return date. In the second task one user indicated a problem with a pull-down menu to select the class because it was displayed partly off the screen. In the third task the users had some difficulties in specifying all the options requested. For the fourth task the only problem seemed to be the long navigation required to reach the page to complete it. Similar issues were reported also for the fifth task in which some users found complications because of a misleading link label.

Regarding the use of our environment for the test, the users generally appreciated its ease of use and lack of intrusiveness. They only found that the automatic creation of the screendump at each page change caused some delay in the interaction. Some users suggested the possibility of creating more detailed user profiles, which can be considered during the analysis phase.

---

[2] `http://mobile.lufthansa.com`

Then, we compared the direct user feedback regarding the mobile application with results obtained with the support of our tool. We thus asked a group with knowledge in usability evaluation to use it in order to identify possible usability problems. We provided them with a short tutorial regarding the tool. We then briefly explained how its various features can be exploited. Then, the evaluators considered each task and the associated user logs in order to identify possible usability problems. We involved six people (three males and three females), average age was 34, all of them with experience in usability evaluation for both desktop and mobile applications, mainly through laboratory user tests or feedback through interviews, questionnaires, and focus groups, thus with little experience in remote usability evaluation, even if they found the approach useful and interesting. They had never used our tool beforehand.

In general, the evaluators found the tool user interface clear and well-structured. They liked the timelines management and the possibility of searching events and pages in them. The evaluators considered how all the users carried out each task through the information provided by the tool. For the first task (Find flight from Paris to London) they soon noticed some long session durations. The occurrence of the second custom semantic event revealed some issues, since it often appeared consistently later than in the optimal session. Likewise, the last custom semantic event associated with the correct flight selection often occurred in an anomalous manner. Some evaluators used the possibility of lining up the timelines according to the pages accessed in the optimal session. This was useful to confirm that in the access to the first two pages the sessions were similar. However, at the page related to the flight selection, the users' behaviour became substantially different. For example, while in the optimal selection the one way option was immediately selected, in some user sessions the user triggered various keyboard events beforehand. This implies that such users made a search without selecting the one way option and then got the request from the application to indicate the date for the return flight. In this case, the involved users either went back to the previous page or stopped their activities for a while. Then, in the end the one way selection event occurred, indicating that they eventually understood the mistake.

In the analysis of the second task (Find information about the lounge in Frankfurt) the evaluators noticed that the users all started from the same page and finished at the right one, but the intermediate pages differed. Indeed, when they searched for the custom semantic event associated with the loading of the page Miles More, they noticed that only one user passed through it. Such differences were well highlighted by the storyboard. The usability issue was that there were two possible ways to accomplish the task, and the shorter was not immediately apparent to the users. Indeed, the path followed in the optimal session requires a link selection (Miles & More) to access information about the lounges, and this was not intuitive.

In the analysis of the third task (Find the flight from Munich to Cracow) the evaluators' analyses focused on the search for the occurrence of the Timetable Page loaded event. Only one user missed that event. Then, the evaluators compared the time taken from when the corresponding page was loaded to when that specific event occurred. For this purpose, some of them combined the search for pages with that for events so as to filter the events to display only the relevant one and align the display according to the flight timetable page in order to compare how long users visited that page. In order to analyse the use of the pull-down menus, the evaluators focused the analysis on the change event in the timetable page and it was thus possible to detect that the use of these elements was not optimal. Some users did not use the filters to narrow the query results and so it took longer to find the requested piece of information. In particular, the usability problem detected was the lack of meaningful labels in the form elements for filtering according to time and flight type.

For the accomplishment of the last task (Find information on Boeing 737-300) the evaluators noticed that the optimal session took about 90 seconds going through six pages while the user sessions usually took about the double of the time. They compared the paths followed in the navigation by the users and found that they sometimes took the wrong direction and then they had to get back to find the right one. The analysis moved to the screendumps of the pages where such deviations occurred and found that users had to select the link “On Board” to access the relevant part of the Web site but that text was not particularly meaningful to understand this.

In general, the use of the environment was satisfying, and improved over time. The comparison of the issues indicated by the end users and those detected in the evaluation through the tool provided similar results.

6 Conclusions and Future Work

We have presented novel solutions for the interactive visualization and analysis of logs associated with users interactions with Web mobile applications in order to better support they remote usability evaluation. We also carried out a user test on an example application that showed that the tool can be used in real remote tests.

Future work will be dedicated to further applying the environment to the evaluation of mobile Web applications, and investigating the integration of the intelligent log analysis with an automatic analysis of the accessed Web pages to check whether they have been implemented according to usability and accessibility guidelines.

References

1. Carta, T., Paternò, F., Santana, W.: Support for Remote Usability Evaluation of Web Mobile Applications. In: ACM SIGDOC, pp. 129–136. ACM Press, Pisa (2011)
2. Hay, B., Wets, G., Vanhoof, K.: Mining Navigation Patterns Using a Sequence Alignment Method. Knowledge and Information Systems 6(2), 150–163 (2004)
3. Nebeling, M., Speicher, M., Norrie, M.C.: W3Touch: Metrics-based Web Content Adaptation for Touch. In: Proceedings CHI 2013, pp. 2311–2320. ACM Press (2013)
4. Paternò, F., Russino, A., Santoro, C.: Remote Evaluation of Mobile Applications. In: Winckler, M., Johnson, H. (eds.) TAMODIA 2007. LNCS, vol. 4849, pp. 155–169. Springer, Heidelberg (2007)
5. Santana, V., Baranauskas, M.: Summarizing observational client-side data to reveal web usage patterns. In: SAC 2010: Proceedings ACM SAC, pp. 1219–1223. ACM Press (2010)
6. Waterson, S., Landay, J.A., Matthews, T.: In the lab and out in the wild: remote web usability testing for mobile devices. In: CHI Extended Abstract 2002, pp.796–797. ACM Press (2002)

Beats Down: Using Heart Rate for Game Interaction in Mobile Settings

Claudia Stockhausen, Justine Smyzek, and Detlef Krömker

Goethe University, Robert-Mayer-Str.10, 60054 Frankfurt, Germany
{stockhausen,smyzek,kroemker}@gdv.cs.uni-frankfurt.de

Abstract. Mobile devices allow integration of different sensors, offering new possibilities for interaction. Integrating heart rate into a mobile game offers several possibilities for enhancing gameplay. In our work we implemented a game prototype on a mobile device with different game modes. Increasing and decreasing heart rate is used for game interaction. The mobile scenario allows involving the environment to influence the heart rate. We conducted a first user experience study for evaluation of the integrated interaction methods in mobile scenarios and conclude with our future work.

Keywords: Physiological Interaction, Mobile Games, Game Interaction.

1 Introduction

While physiological data has a fast growing influence in computer control, as well as in games and evaluation [2], the mobility aspect seems to be neglected. Even though mobile platforms inherit several possibilities to connect and to implement measurement sensors, the aspect of using physiological data for control and interaction within games is yet to be explored in mobile scenarios.

Many games using physiological data adapt game elements or game difficulty based on the measured data. Game mechanics can be enhanced by direct or indirect physiological control [3]. [1] describes a way to integrate different measured data into a commercial computer game engine. During gameplay different game elements and effects were adapted based on the players' physiological signals like e.g. shader effects, volume of the environment and movement speed.

An Adaptation engine, described in [9], uses physiological data for evaluation of user performance. Based on the user performance and several other parameters, like e.g. motor activity, the way of interaction can be adapted

Another field of application for physiological signals which focus more on health aspects are exertion games. In [4] heart rate is used for an exercising biathlon computer game, where the heart rate is used to control the skiing speed. "Heart Burn" [7] uses heart rate scaling to adapt game performance to the fitness level of the player. The mentioned examples are stationary and do not take advantage of interaction with physiological data in mobile scenarios.

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 523–530, 2013.

Only few examples exist using physiological data and taking advantage of the mobile aspect of mobile games. One example is “HeartBeat” [6], which is an outdoor pervasive game for children, in which two teams play against each other. Players wear heart rate sensors and a small device, while trying to capture members of the opposing team. Heart rate is integrated and transmitted to the small device which plays a sound, when an opponent is near and the heart rate of the opponent exceeds a given limit.

Another example is "Zone of Impulse" [5], a simple multiplayer space shooter developed for mobile devices. Galvanic skin response and heart rate are used to conclude on the user state and adapt the game difficulty to allow balancing gameplay. Both games were developed for mobile devices but does not allow to directly control the game by physiological signals or the possibility of integrating/addressing the mobile aspect/using the environment.

The game prototype, we present, uses heart rate as a means of game interaction. Gameplay can be influenced directly by increasing or decreasing the own heart rate. The novelty in our work is the aspect of taking advance of the mobility. This mobility allows the player more flexibility on how the game is played. It offers the possibility to influencing the heart rate by movements and expansive use of the environment to get a deeper gaming experience. One of the important aspects during development was to make sure the connecting devices allow a high mobility range and the minimization of user constraints and boundaries.

2 Prototype

The game is named "Beats Down". In order to reach a broad audience we based the game on a simple principle. The game field (figure 1) consists of 12 fields that flash randomly in red for a short time. To collect points, the gamer has to hit the flashing fields. The goal of the game is to get as many points as possible within a given time frame.

There are different ways to use the heart rate as an input for the game. Heart rate can increase, decrease or be held in a certain zone. Heart rate can even be used for further analysis of heart rate variability, which allows concluding on the mental effort of a person [8]. We decided to focus on the way a player´s heart rate fluctuates, because the fluctuation can be influenced very easy by the player. Therefore, we distinguished between two game modes:

- challenge: elevate gamer heart rate to speed up game play
- relax: lower gamer heart rate to get bonus points

Furthermore one game mode without heart rate has also been integrated. In this simple mode without physiological sensing, the game only consists of hitting the flashing fields without any other input elements.

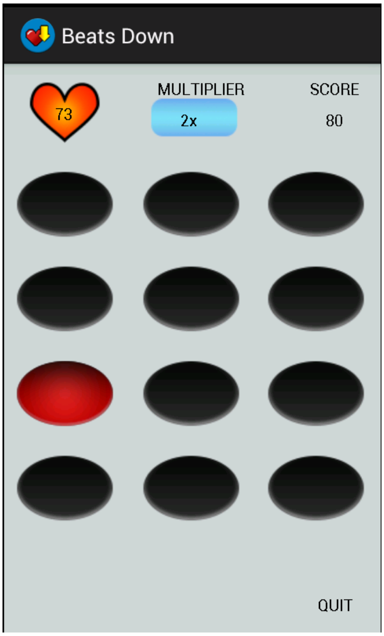


Fig. 1. Layout of Beats Down

At the beginning, the player can choose between these three game modes. If a game mode with heart rate is chosen, the heart rate is displayed at the top of the game field as an additional visual element, next to game speed or point multiplier.

2.1 Challenge Mode

In the challenge mode the speed of the flashing elements is controlled by the gamer´s heart rate. An elevated heart rate correlates with a higher gaming speed. The elements flash up faster, but they also disappear faster. The gamer requires a faster reaction time in order to select the flashing elements, but has increased opportunities of collecting points.

The escalation of the heart rate can be achieved by physical activity or faster breathing. Due to the mobility of the game platform the player is able to engage in various physical activities to increase the heart rate.

2.2 Relax Mode

In the relax mode, the gamer acquires a bonus if the heart rate is decreasing. The more the heart rate is decreased in comparison to the starting heart rate, the more the point multiplier rises, resulting in a higher score for each hit. Each lowering of the heart rate by 10% in comparison to the start value is rewarded with a point multiplier bonus of 1. For example, should the heart rate be lowered by 10% in comparison to the start

Fig. 2. Zephyr HxM heart rate monitor

heart rate, the points will double. To lower the heart rate for the relaxing mode, the gamer can take slow deep breaths or sit/lay down.

2.3 Implementation

The game is developed for Android 4.0.3 devices and optimized for a resolution of 470dp x 320dp. To allow mobile scenarios a sport heart rate monitor is integrated for heart rate measurement. The Zephyr HxM heart rate monitor (figure 2) transmits the data from the chest strap to the game directly via Bluetooth.

3 Preliminary User Test

To evaluate the game's interactivity, participants played all of the different modes of the game. The focus of the test was on the use of heart rate in the game and the aspect of mobility during gameplay. One of the goals was to monitor reactions of the participants upon using the freedom of movement for influencing the heart rate. Furthermore we wanted to find out how fun are these interactions perceived to be.

The user tests started in a usability lab, designed as a living room. Participants had enough space for complex movements and relaxing. Furthermore participants had the possibility to go outside the lab and move freely through the building or campus while playing the game modes with physiological sensing. In both situations, the participants were video recorded and all activities participants choose, were recorded.

3.1 Experimental Procedure

The user test is designed as a three-parameters test per test subject. The three evaluated parameters were: challenging, relaxing and a game mode without heart rate integration. The parameters were set in a randomized order. Each version of the game was played for 2 minutes and after each game session, players had a 2 minute break.

Activities for increasing or decreasing heart rate in the different game modes were recorded.

Before the test, participants had to fill out a demographic questionnaire, answering questions about their previous gaming experience. After each session they had to fill out a questionnaire related to the current game mode and their gaming experience in that particular mode.

3.2 Participants

13 participants took part in the tests (11 male and 2 female). The age of the participants was between 18 and 36 with an average age of 24. Because of health concerns, one person could not play the challenge mode. Therefore we decided to exclude this participant from the analysis of the data. On a scale from 1 to 5 (1=very bad, 5=very good), participants rated their gaming experience on average with 3.9 and their fitness level with 3.5.

4 Results

Analyzing the results of the questionnaire, a Friedman test showed differences in the enjoyment ratings of the participants for the different game modes ($\chi2(2)=16.263$, $p=0.000$). Figure 3 shows the ratings for the different game modes concerning gratification levels experienced through the gameplay. The challenge mode was rated the highest, closely followed by the relax mode.

Investigating the ratings of game enjoyment further, a t-test between challenge and relax mode shows no significant difference ($p=0.1645$). A t-test shows a significant difference between normal mode and the challenge mode ($p=0.00$) and between normal mode and relax mode ($p=0.0025$).

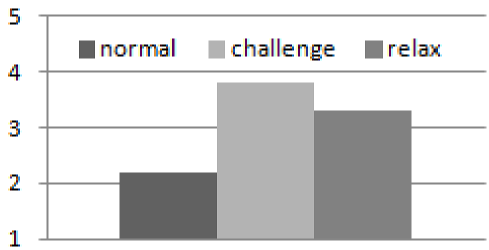


Fig. 3. Results of game enjoyment rating of the different game modes

10 of 13 participants perceived the influence of the heart rate on game play in the challenge mode. While in relax mode, 11 of 13 participants detected the influence of heart rate on their game play.

During the test, participants proved to be highly creative concerning their freedom of movement. Various activities of the participants for increasing or decreasing heart rate were recorded during the test (examples in figure 4). Several test subjects decided to run up and down the stairs while playing the challenge mode. Other activities in the challenge mode were: push-ups, running in the building, knee bends, running on the spot. In the relax mode one participant did yoga to raise the points multiplier. Other relaxing activities were: sitting, breathing exercises, lying down and meditation.

Questioned if they would integrate or use their environment in a mobile game, e.g. running up stairs, the mean rating was 4.07 on a five-point-scale (1=would not integrate/use it, 5= would integrate/use it).

Fig. 4. Participant during relax mode (left) and during challenge mode (right)

During the user study the participants made several comments and recommendations concerning the different game modes and game usability. Most participants asked for a multi-player function, especially for the challenge mode. One participant asked to integrate different options for a bonus in the challenge mode besides game speed.

The participants were overall very interested and enthusiastic towards the implied activities and mobility aspects of the game. Furthermore they would use the offered interactions based on heart rate and enjoyment during gameplay was rated positively; specifically during the challenge mode the results showed an increase in communication and engagement.

5 Conclusion and Future Work

We implemented a first prototype of a mobile game, integrating different forms of interaction based on heart rate. In this game we focused only on two possible interactions but we plan to integrate several different game modes in an upcoming version of the game. A first short study showed that mobility can enhance physiological data integration for interaction into gameplay. Participants showed creativity regarding the use of the environment for interaction. The offered interactions based on heart rate were positively rated in the first user study.

Many possible application areas arise. For example, the mobile aspect of the integration of heart rate for interaction offers the possibility for the development of new exergames, which can be played independently from stationary devices or in combination with location based tasks. Furthermore it would be interesting to investigate further application areas, not only restricted to games.

5.1 Future Game Modes

After analyzing the experiment results, we would like to increase the heart rate degree of interactivity with the gameplay, giving more variation on point rewards and additional game variables. For example rewarding heart rate increases in challenge mode based on percentage with a multiplier for points, similar to the relax mode or other features in future game concepts. While in relax mode incentivize players with visual motivation even for small heart rate reductions and earn time based points for holding such achievements. Furthermore other physiological data (like electrodermal activity) can increase optional information for the challenge and relax mode. This might result in additional modes of physiological feedback and interactivity a player can choose from.

A multiplayer function can increase the challenge level between participants and give us more information about mobility and player interaction. Communication and interaction between player devices can be used for achievement, ranking or feedback purposes.

Additionally, gameplay can be combined with location based tasks, restricting the interaction based on heart rate to specific locations. For example, it can be restricted to a park or a building where you can go for a run or use the stairs.

5.2 Future Studies

To achieve best results and in order to find the optimal game modes, more assessments are necessary. Not only testing with single player modes but also the multiplayer options should be evaluated to verify our current findings. As an additional aspect we would like to attempt using different environments to determine player mobility limits and to expand their range of activities.

References

1. Dekker, A., Champion, E.: Please biofeed the zombies: enhancing the gameplay and display of a horror game using biofeedback. In: Situated Play Proceedings of DiGRA 2007 Conference, pp. 550–558 (2007)
2. Mandryk, R.L.: Physiological Measures for Game Evaluation. In: Isbister, K., Schaffer, N. (eds.) Game Usability. Morgan Kaufmann (2008)

3. Nacke, L., Kalyn, M., Lough, C., Mandryk, R.L.: Biofeedback game design: using direct and indirect physiological control to enhance game interaction. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2011), pp. 103–112. ACM (2011)
4. Nenonen, V., Lindblad, A., Häkkinen, V., Laitinen, T., Jouhtio, M., Hämäläinen, P.: Using heart rate to control an interactive game. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2007), pp. 853–856. ACM (2007)
5. Reitz, K., Stockhausen, C., Krömker, D.: Zone of impulse: physiological data enhanced gaming. In: Proceedings of the 14th International Conference on Human-computer Interaction with Mobile Devices and Services Companion (MobileHCI 2012), pp. 221–224. ACM (2012)
6. Magielse, R., Markopoulos, P.: HeartBeat: an outdoor pervasive game for children. In: Proc. of the 27th International Conference on Human Factors in Computing Systems, pp. 2181–2184. ACM (2009)
7. Stach, T., Graham, T.C., Yim, J., Rhodes, R.E.: Heart rate control of exercise video games. In: Proc. of Graphics Interface 2009, pp. 125–132. Canadian Information Processing Society (2009)
8. Rowe, D.W., Sibert, J., Irwin, D.: Heart rate variability: indicator of user state as an aid to human-computer interaction. In: Karat, C.-M., Lund, A., Coutaz, J., Karat, J. (eds.) Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 1998), pp. 480–487. ACM Press/Addison-Wesley Publishing Co., New York (1998)
9. Duarte, L., Carriço, L.: Coupling interaction and physiological metrics for interaction adaptation. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) INTERACT 2011, Part IV. LNCS, vol. 6949, pp. 499–502. Springer, Heidelberg (2011)

Cultural Congruence and Rating Scale Biases in Homepages

Gitte Lindgaard[1,2], Cathy Dudek[1], and Gerry Chan[1]

[1] Department of Psychology, Carleton University, Ottawa, Canada
[2] Faculty of Design, Swinburne University of Technology, Melbourne, Australia
glindgaard@swin.edu.au, cathy.dudek@gmail.com, gchan@carleton.ca

Abstract. We reanalyzed data from three studies to explore first-impression cultural congruency effects and potential rating scale biases among Canadian and Taiwanese/Chinese participants judging visual appeal of homepages. The objective was to identify variables likely to affect such judgments for future studies in a new research program. Some support was found for both issues and pointers for refinements of future studies were identified.

Keywords: visual appeal, cultural congruence, first impression.

1 Introduction

Interest in cross-cultural user interface design is growing in the HCI community [1]. Studies of the ways companies, governments and universities present themselves on the Web reveal considerable differences between cultures. For example, web sites in collectivist cultures such as Japan and China, use more graphic elements than web sites representing individualistic cultures such as Germany and North America [2]. They differ also in the use of symbols, images, and layouts. Chinese homepages, for example, are often divided into numerous spaces; North American sites are typically arranged around one focal point [3]. However, as culture continues to evolve, the persistence and importance of these cultural differences remain unclear. Some researchers argue in favor of localization of e-commerce web sites even in the current climate of globalization, to ensure visual congruency with different cultures. Since taste changes dynamically, the cost of website localization would be a never-ending task. If people from different cultures also vary in their use of rating scales, the reliability of visual appeal assessment becomes problematic.

Data from the studies included here were analyzed to identify cultural congruency effects and potential rating scale biases in the first impression of homepages from different cultures. If cultural congruency matters, culturally congruent images should attract higher visual appeal ratings than incongruent images. Likewise, if findings show evidence of inter-cultural response biases, then we need to understand how to control or eliminate such biases in future studies. A response bias is a tendency systematically to respond to a range of items on some basis other than what the items

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 531–538, 2013.

were intended to measure [4]. These have been studied extensively in the cross-cultural literature [e.g. 5]. Next, we introduce the notion of cultural congruency, followed by an outline of rating scale biases and then by the analyses of three studies. A general discussion emphasizing future work is then presented.

1.1 Cultural Congruency

Interest in cultural congruency in website design is increasing [6]. Among researchers favoring localized Web design, Noiwan and Norcio [7], for example, note the importance of "empirical investigations on the impacts of cultural factors on interface design" (p. 104). Likewise, [6] argue that "designers should adopt a regional strategy" to reflect user preferences "in various geographical markets" (p. 582). Yet, [8] noted signs of cultural convergence in the way IBM presents itself on the Web, by keeping "the same colors and layout for all localized websites" (p. 1259). This trend is now evident in many global company websites (e.g. HP, Acer). This homogenization, contradicts Simon's [9] assertion that a single "universally appealing global site does not appear feasible" and that companies should "instead create culturally, consumer-specific sites" (p. 32). However, as Gutiérrez and Rogoff [10] note, people participate in dynamic cultural communities in "overlapping ways that change over their lifetimes" (p. 21). The early sociological literature showed that we all adopt different roles in many contexts [11]. Most of our participants belong to a particular student community, to the loosely coupled Internet user community, as well as to other groups. This self-identification with different cultural communities might influence Web design preferences, suggesting that website localization may not be as urgent as localization advocates argue.

1.2 Rating Scale Response Biases

Response biases have been studied in social psychology, personality, and in psychological assessment since the 1940s [e.g. 12], and in cultural studies since the 1960s [13]. Biases have even been found in populations assumed to be relatively homogeneous, such as comparisons of Northern and Southern Italians [14] or Eastern and Western Europeans [15]. Adding studies involving cultures known to differ considerably exacerbates the difficulty of solving the response bias problem. Experimental and statistical control methods have been devised to deal with it, but these are far from always successful [4]. For the purposes of this paper, we reanalyzed data for evidence of three response biases that could affect first-impression judgments of visual appeal. The extremity bias, or 'Extreme Response Style' (ERS), a tendency consistently to select the end points of a rating scale, has been demonstrated in several cross-cultural studies. North Americans are typically found to use the endpoints of rating scales more often than East Asians [e.g. 16]; East Asians tend to select the midpoint of the scale more often [17]. Extreme responses reflect decisiveness and sincerity in some cultures; in others, using the middle of the scale would be seen as attempts to hide one's true feelings. In Asian cultures, cautious responses are seen as modesty; using the extremes of the scale would be in 'poor taste'. These differences reflect variations in emphasis on sincerity versus emphasis on modesty. Both ERSs and the social desirability bias could thus represent socially desirable responses depending on participants' native culture and/or on the degree to which people engage in impression management [18]. Either way, socially desirable responses reflect a desire to 'look good'; they can distort research findings and lead to suboptimal webpage design decisions. The acquiescence bias [19], also called the 'agreement bias' [20], is the tendency to agree with questions regardless of content. Acquiescent responses appear to be more pronounced among people in collectivist cultures. Our study paradigm [21] involved Canadian and Taiwanese or Chinese participants judging the visual appeal of North American or Taiwanese/Chinese homepages in two trials in all three studies presented here. Ratings were made on a 9-point scale (1= least, 9=most appealing). Different stimulus-exposure times were used, but the present analyses only compared average ratings by culture in 50ms trials, collapsed across the homepages in the respective sample.

2 Study 1

Some 50 Chinese and Taiwanese homepages representing the 25 most- and 25 least appealing exemplars of a larger sample of wide-ranging web genres and judged independently were shown to 80 convenience student participants at a Canadian and a Taiwanese University (40 Chinese, 40 Canadian). Visual appeal ratings differed significantly by culture, $t(78)=1.98$, $p=.026$, confirming a cultural congruency effect.

The frequency of ratings of 1 through 9 was calculated by culture as shown in Figure 1 below. A tendency for Canadians to give low ratings more often than the Chinese participants suggests a trend towards an ERS bias for low scores, but none of the independent *t*-tests conducted separately for scores of 1, 2, and 3 was significant. This pattern also suggests acquiescence among the Chinese participants. Alternatively, the Chinese participants may simply have liked these culturally congruent homepages more than Canadians. Further evidence for acquiescence lies in the observation that the Chinese participants gave more ratings at the high end of the scale. Independent-samples *t*-tests performed on ratings of 7, 8 and 9 thus differed significantly on ratings of 7, $t\ (78) = 2.05$, $p < .05$ and 8, $t\ (78) = 2.43$, $p < .05$. Evidently, the ERS was confined to one end of the scale. There was no evidence of social desirability among the Chinese participants (ratings of 4, 5, or 6).

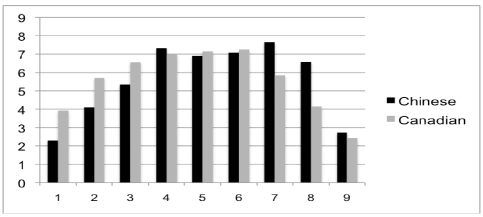


Fig. 1. The frequency of ratings of 1-9 for each of the two cultural groups

3 Study 2

Studies of e-commerce websites have identified significant gender effects, both in within-cultures [22][23], and in between-culture studies [9]. To identify a similar effect in the present paradigm, gender was controlled; participants were Chinese (n=20) and Canadian (n = 20) living in Canada, and the stimuli included only Chinese e-commerce websites. Evidence in the cross-cultural literature suggests that East Asians 'see' more than North Americans in visual tasks [24]. Masuda and Nisbett [25], for example, showed that Japanese participants included information about the context of objects and about relationships among the objects 65% more often than Americans. Similarly, Masuda et al. [26] showed American and Japanese students cartoons depicting a happy, sad, angry, or neutral person surrounded by others expressing the same or a different emotion. They found that Japanese, but not Americans, were influenced by the surrounding people's emotions when judging the focal person's emotion.

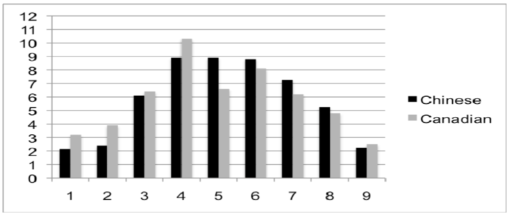


Fig. 2. The frequency of ratings of 1-9 for each of the two cultural groups

There was no cultural congruency effect, but possible reasons for this as well as for the absence of a gender effect are discussed later. The frequency of ratings of 1 through 9 is shown in Figure 2. As in Study 1, Canadians gave slightly more ratings at the low end of the scale (ratings of 1, 2, and 3) than the Chinese participants, again suggesting an ERS bias. However, none of the independent-samples *t*-tests (ratings of 1, 2, 3) were significant. Chinese participants gave slightly more ratings of 5 and 6 than the Canadians, but the *t*-tests for ratings of 4, 5, and 6 revealed a difference only for ratings of 5, $t(38)= 1.691$, $p < .05$, thereby providing some evidence for a social desirability effect. The Chinese participants gave slightly more ratings of 7 and 8 as in Study 1, suggesting slight acquiescence, but the *t*-tests for ratings of 7, 8 or 9 were not significant.

4 Study 3

This study included 50 North American homepages and a sample of Canadian (n = 20) and Taiwanese participants (n = 16) all living in Taiwan. They spoke no English, and they had minimal exposure to Western culture. Higher ratings for Canadians would thus represent cultural congruency. There was a marginally significant effect of culture (p <.058). However, these were actually higher for the Taiwanese than for the

Canadian participants. This finding can be said to represent a ‘reversed’ cultural congruency effect.

The frequency distribution of ratings is shown in Figure 3. Canadians again gave more ratings of 1, 2, and 3 than the Taiwanese participants; the *t*-tests differed for ratings of 2, $t\ (30) = 1.76$, $p < .05$ and 3, $t\ (30) = 1.71$, $p < .05$, again providing some evidence of an ERS bias. The Figure suggests a trend towards social desirability for the Taiwanese participants. However, the *t*-tests for ratings of 4, 5, and 6 differed only for ratings of 5, $t\ (30) = 2.12$, $p < .05$, thus again providing partial support for social desirability. Starting with ratings of 4, the Figure shows a trend towards Taiwanese acquiescence except for ratings of 9. The t-tests (ratings 7, 8, 9) were not significant.

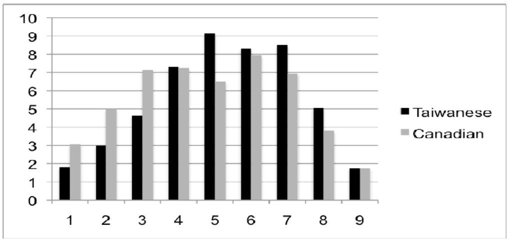


Fig. 3. The frequency of ratings 1-9 for each of the two cultural participant-groups

5 General Discussion, Ongoing and Future Studies

A cultural congruency effect was found in Study 1 exposing Chinese and Taiwanese homepages representing a wide variety of websites. No such effect was found in Study 2, which used only Chinese B2C e-commerce sites. At first glance, this seems strange. However, an anecdotal comparison of ratings by culture and gender revealed an interesting pattern. Ratings for Chinese females were invariably somewhat higher than for Canadian females, but this was reversed for males where Canadian males invariably rated the homepages somewhat higher than Chinese males. This would eliminate any potential effects of culture. Other studies have found consistent and significant gender differences both in male and female-produced websites, and in male and female preferences for websites designed by members of their own gender. Moss and her colleagues, for example, identified 23 variables related to visual appeal, navigation and use of language of which 13 were found to differ significantly between male- and female-produced websites (Moss & Gunn, 2009). Thus, females were more likely than males to use rounded than straight shapes, to use more colors, and to avoid a horizontal layout in their personal homepages. Participants were blind to the designer’s gender, but preference data showed that both genders preferred websites designed by members of their own gender. Since the number of males in the IT industry by far exceeds the number of females, it is possible that this factor influenced our findings. We have no way of knowing the gender of designers producing the

Table 1. Summary of rating scale biases

Study	Homepages	ERS	Soc. Desirability	Acquiescence
1	Taiwanese/Chinese	Trend low end	No	Yes, some
2	Chinese	Trend, low end	Yes, some	Trend only
3	North American	Yes, low end	Yes, some	Trend only

homepages used in our studies, but we are currently running a study in which both the designers' and the participants' gender is carefully controlled. Findings will be reported at INTERACT.

The reverse cultural congruency effect on North American homepages could represent a novelty effect among the Taiwanese participants who had almost no exposure to English-language websites. We are planning to investigate this possibility with a series of studies controlling more carefully for both gender and previous exposure to local websites in different cultures. As in Moss and her colleagues' studies, we will use samples of personal websites, as these reflect the way individuals wish to present themselves on the Web. That is, they are not bound by any corporate design restrictions, team-based decisions, or other constraints.

Concerning rating scale biases, Table 1 summarizes the trends in the above studies. Although the three studies provided some evidence in support for all three biases, these are neither robust nor systematic. One possible explanation is that, although the homepages represented the most and least visually appealing pages from larger samples, they may not have been sufficiently polarized to yield clearer results. Given the apparent importance of Web designers' gender, careful control for this should be more likely to yield a more clear-cut insight into potential rating scale biases. Rating scale biases reported in the literature are typically found in studies seeking degree of agreement with textual statements or judgments of latent variables such as the emotion of a focal person. The biases may not be as easily identified in studies requiring quite different kinds of judgments, for example, of visual appeal. Still, with the growing interest in inter-cultural studies and the trend towards globalization of websites, it is important to pursue this possibility further. We therefore include analysis for such biases in our current and future studies. In conclusion, it appears that cultural congruency, rating scale biases, and gender differences do exist in the production and assessment of websites. All of these are taken into consideration and controlled for in our studies currently underway, to pinpoint exactly where these effects are greatest.

Acknowledgements. The research was supported by a Canadian National Science & Engineering Research Council (NSERC)/Cognos Industry Research Chair grant, IRCSA 234087-05.

References

1. Bonnardel, N., Piolat, A., Le Bigot, L.: The impact of colour on website appeal and users' cognitive processes. Displays 32, 69–80 (2011)
2. Segev, E., Ahituv, N., Barzilai-Nahon, K.: Mapping diversities and tracking trends of cultural homogeneity/heterogeneity in cyberspace. Journal of Computer-Mediated Communication 12, 1269–1297 (2007)

3. Schmid-Isler, S.: The language of digital genres – a semiotic investigation of style and iconology on the world wide web. In: Proceedings of the 33rd. Hawaii International Conference on System Sciences, HICSS 2000 (2000)
4. Paulhus, D.L.: Measurement and control of response bias. In: Robinson, J.P., Shaver, P.R., Wrightsman, L.S. (eds.) Measures of Personality and Social Psychological Attitudes, ch. 2, pp. 17–59. Academic Press Inc., San Diego (1991)
5. Nisbett, R.E., Choi, I., Peng, K., Norenzayan, A.: Culture and systems of thought: Holistic versus analytic cognition. Psychological Review 108(2), 291–311 (2001)
6. Kim, H., Coyle, J.R., Gould, S.J.: Collectivist and individualist influences on website design in South Korea and the U.S.: A cross-cultural content analysis. Journal of Computer-mediated Communication 14, 581–601 (2009)
7. Noiwan, J., Norcio, A.F.: Cultural differences on attention and perceived usability: Investigating color combinations of animated graphics. International Journal Human-computer Studies 64, 103–122 (2006)
8. Kondratova, I., Goldfarb, I.: Cultural visual interface design. In: Proceedings of EdMedia 2005, World Conference on Educational Multimedia, Hypermedia & Telecommunications, Montréal, Canada, pp. 1255–1262 (2005)
9. Simon, S.J.: The impact of culture and gender on web sites: An empirical study. The DATA BASE for Advances in Information Systems 32(1), 18–37 (2001)
10. Gutiérrez, K.D., Rogoff, B.: Cultural ways of learning: Individual traits or repertoires of practice. Educational Researcher 32(5), 19–25 (2003)
11. Merton, R.K.: The Role-Set: Problems in Sociological Theory. British Journal of Sociology 8(2), 106–120 (1957)
12. Cronbach, L.J.: Response sets and test validity. Educational and Psychological Measurement 6, 475–494 (1946)
13. Triandis, H.C., Triandis, L.M.: A cross-cultural study of social distance. Psychological Monographs 76, 1–21 (1962)
14. Knight, K.N., Nisbett, R.E.: Culture, class, and cognition: Evidence from Italy. Journal of Cognition and Culture 7, 283–291 (2007)
15. Varnum, M.E.W., Grossman, I., Kitayama, S., Nisbett, R.E.: The origin of cultural differences in cognition: The social orientation hypothesis. Current Directions in Psychological Science 19(1), 9–13 (2010)
16. Jenkins, L.J., Yang, Y.J., Goh, J., Hong, Y.Y., Park, D.C.: Cultural differences in the lateral occipital complex while viewing incongruent scenes. SCAN 5, 236–241 (2010)
17. Lee, J.W., Jones, P.S., Mineyama, Y., Zhang, X.E.: Cultural differences sin responses to a Likert scale. Research in Nursing & Health 25, 295–306 (2002)
18. Goffman, E.: The presentation of self in everyday life. Penguin books Ltd., Ringwood (1959)
19. Smith, P.B.: Acquiescent response bias as an aspect of cultural communication style. Journal of Cross-cultural Psychology 25(1), 50–61 (2004)
20. Johnson, T., Kulesa, P., Cho, Y.I., Shavitt, S.: The relation between culture and response styles: Evidence from 19 countries. Journal of Cross-cultural Psychology 36(2), 264–277 (2005)
21. Lindgaard, G., Dudek, C., Sen, D., Sumegi, L., Noonan, P.: An exploration of the relations between visual appeal, trustworthiness, and perceived usability of homepages. Transactions on Human-computer Interactions (TOCHI) 18(1), 1–30 (2011)
22. Cyr, D., Bonanni, C.: Gender and website design in e-business. International Journal of Electronic Business 3(6), 565–582 (2005)

23. Moss, G.A., Gunn, R.W.: Gender differences in website production and preference aesthetics: preliminary implications for ICT education and beyond. Behaviour & Information Technology 28(5), 447–460 (2009)
24. Hamamura, T.: Are cultures becoming individualistic? A cross-temporal comparison of individualism-collectivism in the United Stages and Japan. Personality and Social Psychological Review 16(1), 3–24 (2012)
25. Masuda, T., Nisbett, R.E.: Attending holistically versus analytically: comparing the context sensitivity of Japanese and Americans. Journal of Personality and Social Psychology 81, 922–934 (2001)
26. Masuda, T., Gonzalez, R., Kwan, L., Nisbett, R.E.: Culture and aesthetic preference: Comparing attention to context of East Asians and Americans. Personality and Social Psychological Bulletin 34, 1260–1275 (2008)

Diverse Ecologies – Interdisciplinary Development for Cultural Education

Michael Heidt[1], Kalja Kanellopoulos[1], Linda Pfeiffer[2], and Paul Rosenthal[2]

[1] Research Training Group 'crossWorlds', Chemnitz University of Technology, Thüringer Weg 5, 09126 Chemnitz, Germany
[2] Visual Computing Laboratory, Chemnitz University of Technology, Straße der Nationen 62, 09111 Chemnitz, Germany
{michael.heidt,paul.rosenthal}@informatik.tu-chemnitz.de, kalja.kanellopoulos@phil.tu-chemnitz.de, linda.pfeiffer@s2011.tu-chemnitz.de

Abstract. We present a case study outlining development efforts towards an interface ecology to be deployed in museums. We argue that the problem at hand calls for a highly interdisciplinary design process. Furthermore, system design in the domain of cultural education poses a unique set of challenges. At the same time few existing design methodologies are suitable for addressing this special environment of system design. We outline a set of tentative methodological elements aimed at informing adequate interdisciplinary development processes. The discussion is embedded into a critique of existing methodologies while being orientated towards inviting critique itself. The guiding insight steering our methodological developments is that fundamental differences between project participants and other stakeholders should be construed as assets. Rather than trying to integrate them or covering them up, the dynamic friction between differing viewpoints can be rendered productive by means of poietic practices.

Keywords: interdisciplinarity, museum informatics, design for cultural experience.

1 Introduction

Devices geared towards product development aim at producing artefacts amenable to concepts such as efficiency, effectiveness and usability. Cultural contexts usually require a different development perspective. Most interfaces constructed for exhibition spaces do not aim at enabling the user to fulfil specific and predetermined tasks as quickly as possible, but offer the possibility to explore and experience the exhibited artefacts or the narratives connected to them. They have to enrich visitor experience and at the same time need to refer to the context they are placed in. Therefore, the actual situation of deployment and use gains enormous importance. It consequently needs to be accounted for within the design process at an early stage. What is to be observed on the contrary is that technological developments for museums are often decontextualised during the process of design. They are frequently outsourced to

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 539–546, 2013.

external production companies, while cooperation between these companies and the museums during the design process remains rudimentary. This leads to the fact, that the highly deployment specific actual contexts of use for these technologies often are neglected. Consequently, many of the developed technologies are not well received by their end users, the museum visitors. In order to overcome the mentioned problems, we state the necessity of realigning technological systems with the cultural contexts they are to be operated in.

2 Project Context

The project described here is tasked with developing interactive installations and mobile applications for the domain of cultural education within museums. It is situated within a collective of researchers with diverse backgrounds. These range from informatics and engineering over psychology and sociology to philosophy, rhetoric, and literary studies. Installations are developed in close collaboration with a local museum of industry, while collaborations with other museums exist as well. Among the installations developed are tangible user interfaces, mobile applications and multitouch tables. The design processes are informed by an exploratory pre-study and a concurrent ethnography. Some of the results of the pre-study and their implications for design will be outlined in the following.

Concerning new media and interfaces in museums and exhibitions the discursive focus often lies on learning [1]. What can be observed empirically on the contrary, is the importance of the visitor *experience* during a museum visit. This experience is influenced by several factors. Not only is the museum an environment for cultural education, but also one for strolling, interaction and communication with others. The museum environment connects visitors to those far in time and space: to past events, persons deceased, artefacts, concepts and narratives rendered strange by the passing of time. At the same time visitors have to interact and connect with those close to them in space and time. This includes interaction with other visitors as well as the interactivity with exhibits and technical artefacts. It is this dynamic interplay of times, spaces, persons and artefacts that forms the unique ensemble of situations that constitutes the museum experience itself. It is characterised by the contrast of an immersive present environment with the invoked feelings of something far removed. Consequently, the aim of our design process is not to transform the experience of the museum visit in a spectacular way, but to support it gently and on a playful basis.

In the subsequent phase of the project, prototyping techniques were deployed in order to generate ideas for improvement of existing systems. The overarching narrative was that of a museum-visit. It exhibits a three-part structure: preparation, actual visit and recapitulation. It has proven essential to develop overarching scenarios in order to contain specialists' affinities to disregard context specifics and to selectively optimise their respective system components regarding technological impressiveness and feasibility.

Among the prototypes developed were:

1. A recommender engine, trying to foster communication by pointing users towards fellow visitors, possessing relevant knowledge.
2. A multitouch-table installation, allowing users to engage in playful interactions with exhibit related information.
3. An interactive role-playing device. Optical recognition of props present in the museum is used to present users with relevant narratives.

The development of the first set of prototypes and scenarios was followed by interdisciplinary discussions. Subsequently and regarding the outcome of this exchange, the attempt was made to break up monolithic deployments into their conceptual and material parts.

To this end usage scenarios for individual prototypes were discussed. Considering the observations made by social scientists in the field, usage scenarios were elaborated into mini-narratives accounting for situational factors within the deployment context. These mini-narratives now can be composed in a non-linear fashion, thereby allowing for exploring possible patterns of system usage.

For example, the analysis unveiled that though prototypes i) and ii) had been developed with two distinct sets of goals in mind, they share a functional characteristic. Both facilitate communication between visitors. The recommender system does this by virtue of direct proposals for communication. The table installation reduces communicative inhibitions by allowing visitors to perceive each other during technologically mediated interactions. The analysis thus allowed for a more conscious and strategic utilisation of the communicative incentives provided.

In a next step individual elements will be orchestrated into new functional ensembles according to the requirements of the museum ecology.

3 Implications of Interdisciplinarity

The value of technological virtuosity is not to produce ever more spectacular displays of its own prowess. It is twofold in nature. First, it allows more degrees of freedom when designing technology. As social practice is diverse, reproduced within a web of disparate sites of practice, the technical possibilities for cultural embedding of technology likewise are diverse and disparate. It is an informatician's job to recognize and exploit these possibilities. This requires a vast level of mastery of technological skills – as well as insights in the social and cultural structures the developed technology will be deployed in. Therein lies one of the main reasons for interdisciplinary cooperation with social scientists in system development. Collaborations of this type are based on a high level of communication between heterogeneous actors – apart from information professionals and social scientists, system users and other professionals within the technological field itself have to be included as well.

The key in facilitating communication in these kinds of interdisciplinary working contexts lies in the realisation that system development too contains theory building aspects. Whenever confronted with a new domain, developers usually develop informal theories describing how this domain is structured. They do not do so by means of ethnographic analysis and generally do not apply social theory. At this point, the theory building process is usually implicit and remains opaque to other project participants. This does not mean, however, that the theories themselves are unstructured or inherently informal. On the contrary, system designers with a background in engineering tend to construe social practices observed with the same formal concepts used to understand technical components. This creates potentials for misconstruction of the social. While diverse perspectives and ambiguities can have catastrophic effects within the sphere of the technical, understanding of social practice necessitates embracing diversity, heterogeneity, multi-perspectivity – and calls for acceptance of its inherent ambiguities.

The relationship between design, social science and informatics has been discussed extensively within the HCI community [2–4]. Following the authors cited, we acknowledge that there is tension inherent in the cooperation of engineers and social scientists. We argue, however, that this tension can be made productive by employment of adequate methodologies. Therefore, our approach does not try to integrate differing perspectives within one conceptual framework. We rather try to create sensitivity of project participants towards the perspectives of the others, as outlined in Fig. 1. We do not expect to achieve fundamental perspective transformations, but aim at repeated chiastic crossing of boundaries.

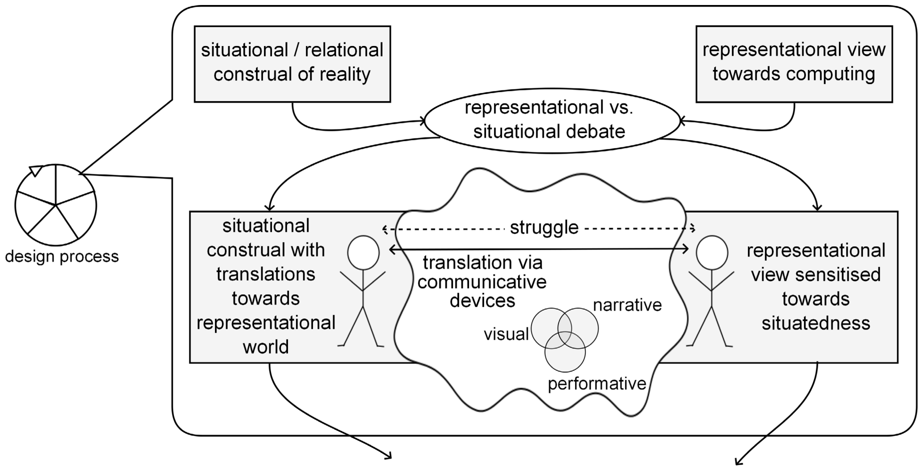


Fig. 1. Differing perspectives are made productive by means of translations and mutual sensitivity

While it is not necessary for social scientists to fully grasp technical intricacies and engineers do not need to follow involved discussions of social theory or methodologies to their conceptual origins, it is paramount that project participants develop a feeling of 'how they are different'. Inside the disciplinary terrain the same demand needs to be made, as researchers from the same discipline can differ broadly concerning research paradigms, methods and methodologies as well as favoured theoretical approaches. We developed a set of communicative devices in order to highlight these differences and make them tractable within the design process. These are detailed in the following subsection. Among the techniques employed are different forms of diagrams as well as other visual aids and narrative devices.

Communicative Devices

In order to aid the communicative process within the interdisciplinary setting, different forms of analytical description were employed. These ranged from formal-technical diagrams employing mathematical language over pictures and photos to informally presented narratives. It has proven both necessary as well as beneficial to encourage participants to discuss their work at an early stage. Furthermore, scientists had to explain their different backgrounds to each other, including previous work and methods employed. A multitude of mutual misconceptions could thus be caught at an early stage – while others may still be present.

Additionally, it has to be stressed that both spatial proximity of researchers from diverse fields, as well as their spatial distance to the surrounding everyday university processes, has proven to be an invaluable asset in producing ideas. As so often is the case, communal trips to the canteen, spontaneous doodle sessions during coffee breaks and seemingly random prototyping sessions in the printer-room have dwarfed official and semi-official design workshop appointments regarding creative power.

However, this should not point to the conclusion that informal talk is enough, rendering formal appointments obsolete. On the contrary, regular meetings have proven to be an integral and invaluable constituent of the communicative ecosystem. They provided helpful incentives to translate work into a language accessible to the other. Frequently, forcing to take these fresh perspectives resulted in the emergence of new ideas and a firmer grasp of the tacit assumptions underlying the own work process. Consequently, a practice similar to that of SCRUM [5] was adopted: Participants were encouraged to explain what they have been working on since the last regular meeting. Encouraging participants to pose questions to each other also proved to be helpful. Within these discussions it proved helpful to develop a certain discipline: Before presenting an argument, dialogue partners had to adequately sum up the position of their predecessor. Furthermore, conducting clustering sessions where participants had to cluster thoughts and concepts produced by researchers from another discipline furthered mutual understanding. A set of visual devices was also employed, these are detailed further in an accompanying publication [6].

The most essential communicative device was that of prototyping. Prototyping provides means to overcome communicative boundaries by employing the integrating power of the material. Thereby, prototypes produced create unique communicative opportunities. Problems and perspectives can be discussed across disciplinary languages and vocabularies with respect to artefacts, which literally are graspable. They allow researchers and developers to derive descriptions from concrete material objects. At the same time, they can be discussed as material correlates of ideas. In the outlined project prototypes were cut out of paper and cardboard, assembled using everyday objects forming structures bearing similarity to an assemblage or made from dough or lego bricks.

However, prototyping can also gloss over antagonistic perspectives: Since the prototype already has been built, utterances can easily be misunderstood by presupposing the own construal of the artefact at hand. Frequently, the opportunity to shatter and question presuppositions is missed during the process, especially if it is geared towards rapid development. Therefore, it proved imperative to embed the prototyping process into the network of communicative support detailed above. Every prototype developed has to be discussed prior to inception, during the respective prototyping session as well as afterwards during a regular meeting.

4 Related Work

At this point it is important to address the inherent contradictions of user-centered design (UCD). Central to the UCD process usually is an understanding of users' tasks and goals. However, in most domains the users' tasks will not be their own. They are determined by organisational and institutional settings. Actual users' wishes and desires will entertain a dialectical tension with those environmental requirements. Furthermore, we highlight the role of UCD as being rooted in a rhetorical situation. UCD is not employed by uninterested third parties but by persons and organisations wishing to achieve certain goals.

Our treatment of system design is informed by the notion that concepts have politics [7] and that artefacts facilitate social practices. System development cannot merely be considered as an instrumental process. It is a site of struggle, where different interests and views of the world engage with each other. Our methodologies try to account for these productive antagonisms while trying to enable productive discussion and struggle within the embodied practices themselves. Our goals are distinct from UCD in that we do not aim at merely identifying user expectations or desires. What we envision is not exclusive development for the user but a critical dialogue between project stakeholders *as well as* end users.

Embodied Interaction developed by Paul Dourish [8] proposes a blend of methods from social-sciences and humanities. He builds on a phenomenological understanding of the realm computing systems are part of. Drawing on Lucy Suchman's research [9], he develops a processual conception of reality in conjunction with a situational multi-level conception of meaning. The key concept is that of embodiment, reminding designers how each artefact always is used in the context of practice and that practice in turn takes part in a meaning making process. Dourish goes on to develop a set of practical design guidelines. We found the argumentation to be cogent and the specification of practical guidelines refreshing. Embodied Interaction provides a site for common understanding between social-scientists and informaticians, due to its roots both in anthropology as well as informatics. It has consequently informed our design process and proved to be one of the most beneficial conceptual toolsets. However, it focusses more on the analysis of the socio-ontological status of HCI artefacts while temporarily neglecting production processes. Consequently, we found that it needs to be combined with further methodological tools in order to realise its inner potential.

Participatory Design aims at communicative integration of all stakeholders potentially affected by an artefact [10]. It centres on democratic discussion and deliberative processes. Although our own approach is informed by Participatory Design, we do not conceive of our design project as an instance of public service. The design team follows convictions not identical with that of potential users. Consequently, we do not try to simulate a consensus. Rather, we try to articulate the privileged nature as well as the limitations of our own position as scientists, developers, designers.

Value-Sensitive Design [11–14] describes itself as an alternative paradigm to participatory approaches. Whereas Participatory Design highlights the need for valuing every participant's opinion, here adhering to human values is given preference. Value Sensitive approaches emphasise the need to acknowledge trade-offs between conflicting values. Consequently, the design process is aimed at enabling reasonable ethically informed decisions, while stating that these must remain publicly accountable.

Our design process is informed by extensive discussions of ethical implications. In contrast to Value Sensitive Design, we do not try to create a deliberative space integrating each and every stakeholder. This first of all is not feasible from an organisational perspective within our project parameters. While we do try to involve potential system users early in the process, this is not done in order to reach a consensus. On the contrary, we think it is imperative to make our own goals very clear and embody them into artefacts. These should be contestable in a wider realm of discussion. In effect, we view our design actions as being accountable to others, while not trying to speak from a position of consensus.

Critical Technical Practice as initially put forward by Phil Agre constituted itself as an effort to reform the field of Artificial Intelligence [15–17] . Within the works cited, Agre argues for a situational approach towards AI problems, trying to highlight shortcomings inhering in the idea of universal cognition. Our research efforts align themselves with some of the basic stances put forward by Agre. Namely, we try to reflect on the untested assumptions underlying practices of technology design.

Critical Technical Practice has since informed a multitude of concepts within the field of HCI. One of these influences is Reflective HCI [18], a joint effort of theorists affiliated with Embodied Interaction, AI centred Critical Technical Practice and other paradigms within the HCI community.

5 Conclusion

We have described an interdisciplinary development project in the domain of museum education. Highlighting the specificity of cultural contexts, we argued for the necessity of customising existing design methods. Viewing the diversity of differing perspectives as assets rather than problems, we have described communicative devices to inform interdisciplinary development. Consequently, our process aims for repeated chiastic crossing of boundaries in order to sensitise participants for the otherness of the other. We do not expect to achieve fundamental perspective transformations on a short or medium timescale.

References

1. Reussner, E.M., Schwan, S., Zahn, C.: New Technologies for Learning in Museums: An Interdisciplinary Research Project. In: Trant, J., Bearman, D. (eds.) International Cultural Heritage Informatics Meeting (ICHIM 2007): Proceedings. Archives & Museum Informatics, Toronto (2007)
2. Mackay, W., Fayard, A.: HCI, natural science and design: a framework for triangulation across disciplines. In: Proceedings of the 2nd Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques, pp. 223–234 (1997)
3. Hughes, J.A., Brien, J.O., Rodden, T., Plc, N.B.: Designing with Ethnography: A Presentation Framework for Design, pp. 147–158
4. Hughes, J., King, V., Rodden, T., Andersen, H.: Moving out from the control room: ethnography in system design Moving out from the control room: ethnography in system design (1994)
5. Schwaber, K., Beedle, M.: Agile software development with Scrum. Prentice Hall PTR, Upper Saddle River (2002)
6. Heidt, M.: Examining Interdisciplinary Prototyping in the Context of Cultural Communication. In: HCI International 2013 (in press, 2013)
7. Suchman, L.: Do Categories Have Politics? The language/action perspective reconsidered, pp. 177–190 (1994)
8. Dourish, P.: Where the action is: the foundations of embodied interaction. The MIT Press (2004)
9. Suchman, L.A.: Plans and situated actions: the problem of human-machine communication. Cambridge University Press, New York (1987)
10. Muller, M., Kuhn, S.: Participatory design. Communications of the ACM (1993)
11. Friedman, B., Borning, A.: Value Sensitive Design as a Pattern: Examples from Informed Consent in Web Browsers and from Urban Simulation. Design (1997)
12. Friedman, B., Kahn, P.H., Hall, G.: A Value-Sensitive Design Approach to Augmented Reality. Design, 163–164 (2000)
13. Louise, J., Davis, N.: Value Sensitive Design of Interactions with UrbanSim Indicators. Design (2006)
14. Borning, A., Muller, M.: Next steps for value sensitive design. In: Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems - CHI 2012, p. 1125 (2012)
15. Agre, P.: Toward a Critical Technical Practice: Lessons Learned in Trying to Reform Al. Social Science, Technical Systems and Cooperative Work: Beyond the Great Divide, 1–17 (1997)
16. Agre, P.E.: The soul gained and lost: artificial intelligence as a philosophical project. Stanford Hum. Rev. 4, 1–19 (1995)
17. Agre, P.E.: Computation and Human Experience. Cambridge University Press (1997)
18. Dourish, P., Finlay, J., Sengers, P., Wright, P.: Reflective HCI: Towards a Critical Technical Practice (2004)

Evaluation of PhonAge: An Adapted Smartphone Interface for Elderly People

Farah Arab, Yasir Malik, and Bessam Abdulrazak

University of Sherbrooke, Département d'informatique
2500 boulevard de l'Université, Sherbrooke, Québec, Canada J1K 2R1
{farah.arab,yasir.malik,bessam.abdulrazak}@Usherbrooke.ca

Abstract. Smartphones can play a significant role in maintaining decent Quality of Life for elderly people. Key factor to Smartphones usage success among elderly people depends on the accessibility of phone interface. In this paper, we present preliminary evaluation results of our PhonAge, an accessible and adaptable interface for Smartphones that is customized to the elderly profile. The results of the evaluation show adequacy of the interface to elderly needs. The evaluation feedbacks also helped in improving PhonAge interface design.

Keywords: Smartphone, Mobile Phone Interface, Accessibility, Evaluation, Aging.

1 Introduction

Smartphones can play a significant role in maintaining decent Quality of Life (QoL) for elderly people. It can promote independent living, active aging, and active socialization, while ensuring safety. The aging process is associated with progressive degradation of sensory, physical and/or cognitive abilities. The related disabilities deprive elderly people from independent living and maintaining social interaction [6], as a consequence they feel isolated and are transferred to special care facilities.

Recently, Smartphones have gained attention among elderly people and many of them are starting to use them for various purposes (e.g. as a reminder) [7]. Numerous services available on Smartphones can assist elderly people (e.g. navigation, interactive diary, activity reminder, speed dial, emergency service), however these solutions are not much welcomed in elderly community, either, because of accessibility issues, usability, elderly resistance and rejection toward technology, complex and tedious perception toward technology, etc. [3, 8]. At present, the mobile industry is targeting young and a tech-savvy population [5] and does not always take into account the special needs and expectations of the elderly population when designing services or phone devices. The main difficulties in adapting Smartphones for elderly people are limited or lack of personalization, complexity of mobile user interface (mUI) or a navigation menu, quantity of information arranged on screens, number of functions, language, etc. Consequently, mobile technology becomes more and more complex and elderly people are confronted with devices and services that are not adapted to their needs. We present in this paper a preliminary evaluation study

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 547–554, 2013.

that is conducted to test the usability of our solution named PhonAge [1], an accessible and adaptable solution for Smartphones that can host diverse useful services to elderly people. The results and feedback from this evaluation illustrates the adequacy of PhonAge to elderly needs, and helped in addressing the shortcomings of PhonAge interface.

The rest of the paper is organized as follow. In the next section, we present an overview of PhonAge, in section 3 we present the evaluation of PhonAge along with discussion on results in section 4, and finally, in section 5 we conclude our work with direction on our future work.

2 Overview of PhonAge

PhonAge is an attempt to promote the use of smartphones among elderly people. In our experience while designing mHealthcare and social services on Smartphones for elderly people, we found that accessibility of phones is the major obstacle towards acceptance, usability and continuity of use [1]. To address such issues, we designed and evaluated PhonAge, an accessible and adaptable solution for Smartphones that is customized to the elderly profile and can host diverse services to assist them in their activity of daily living (ADL) [1]. PhonAge main interface/screen (Figure 1-a) is divided in three sections, namely **(1)** a spatiotemporal assistance section (Top section) to continuously provide elderly people with spatiotemporal contextual information (date, time and weather), **(2)** the Main activity and service area section to provide accessible phone feature and services to assist elderly and **(3)** an emergency section to ensure safety of elderly people while moving. PhonAge emergency button allows elderly people to alert someone in case of emergency (e.g. health problem, attack, wandering) according a predefined context-aware emergency protocol (based on the context-awareness framework presented in [2]). It also provides a configuration options that allow elderly people/caregiver/family to adapt Smartphones according to elderly needs, change the text and icon size, choose among the icons from gallery, change the order of icons, etc. (detailed presentation of PhonAge is available in [1]).

3 Evaluation

To study the usability and accessibility factors of PhonAge, we conducted a first stage evaluation with elderly people. The study aims at highlighting the compliance of PhonAge's mUI to the needs of elderly people and identifies usage difficulties. Following, we present evaluation method and preliminary results.

3.1 Method and Data Collection

To evaluate the mUI of PhonAge, we compiled a semi-structured interview supported by a questionnaire of about 45-60 minutes composed of 27 questions divided in three parts, and a direct observation of the PhonAge usage. The interview is designed to be conducted in person (by examiner) and the feedback and remarks to be recorded by the examiner. The questionnaire was pre-tested with one elderly people of age 65. Following we describe three parts of questionnaire:

(a) (b)

Fig. 1. PhonAge Main Screen (a) and Icon Presentation (b)

Part. 1. Gathering user information: In this part, we ask questions related to participants profile, such as age, education level, living environment (independent or nursing home), etc. The participants were asked to rate each item using a 5-point Likert scale.

Part. 2. Gathering details of outdoor ADLs: In this part, we ask questions about outdoor ADLs (i.e. purpose, frequency and organization of activities), kind of support for outdoor activities (e.g. mobility equipment device), use of mobile phone purpose (e.g. call, messages, reminder), frequency of mobile phone use, type of phones used (e.g. touch screen, dailpad).

Part. 3. Accessibility and Usability Test: In this part, we ask questions about understandability, usage, utility and satisfaction of the PhonAge interface.

- Icon Understandability Test is run to evaluate the meaning that PhonAge icons give (similar to [4]). First, participants are asked to give the meaning of the 12 icons proposed in PhonAge (figure 1-b). Second, we present the icons with their respective labels and ask the following question: "*According to you which information do you consult if you touch the next icons*". The order of the icons is modified to avoid inferring with any meaning.
- Use Test Method is run to evaluate the usability of PhonAge. The evaluation is designed to assess difficulties in use, utility of the navigation arrows provided (to switch between screens), and utility of functions (e.g. emergency, organization of icons/functionality on main screen). Participants are asked to navigate through phone functions and perform four defined tasks i.e. ***call, search contact address, search email and locate the camera function***.

Moreover, participants are asked to evaluate the PhonAge interface in terms of utility, facility of use and their preferences (i.e. interface structure, color of the wallpaper, services usefulness). We ask questions like "*According to you, is it easy to? According to you, it is useful for outside activities to? According to you, the background is?*"

4 Results

The results presented in this paper are those conducted with twenty elderly persons aged 60 to 84 (Mean=70 years, standard deviation=8, 2, 13 females and 7 males) who participated in this study. Following, we present the results.

4.1 Activity of Daily Living

In our study, most of the participants (19/20) live at home except one person lives in a retirement home. 11/20 participants live with family, 5/20 live in couple and 4/20 by themselves. The results show that the main outings are grocery (14/20), daily walk (10/20), visit relatives (10/20), leisure (8/20), medical appointments (5/20), administrative activities (2/20) and volunteering (1/20). 10/20 participants prepare independently their outings, 8/20 by writing a task list, 3/20 plan their trips (e.g. bus schedules), and 3/20 assemble in advance documents for administrative activities. 7/20 has been at least once victim of abuse outside. Half of the participants (10/20) also don't feel secure outside particularly because of the risk of street attacks or falls, vulnerability feelings and frailness. However, this insecure feeling decreases when someone accompanies them. Results show also that 11/20 participants have a mobile phone that they use either to call, to be contacted or as an event reminder (e.g. birthday or medical appointments). 3/20 participants also use the phone to send a message or take a photo, and 5/20 participants consult their directory (occasionally). Figure 2 illustrates the results of Use of mobile phones by elderly people.

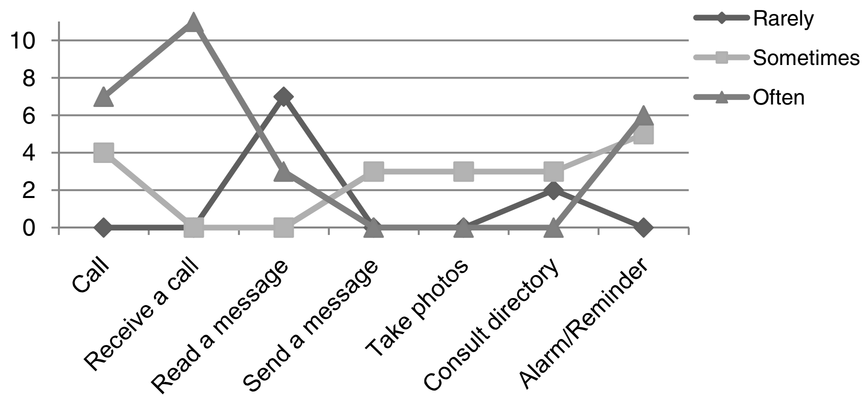


Fig. 2. Use of mobile phones by elderly people

4.2 Icons Understandability Test

Evaluation of the icon understandability showed, that common icons (e.g. phone, emergency) were easily understandable (meaning and purpose) by all participants (20/20), even without a label (Figure 3 illustrates the results of the understandability of PhonAge icons). Labeling the icons helped participants to understand the navigation and social service icons. However, none was able to understand the following two icons: "shopping" for shopping list and "home phone number" on contact screen (Figure 5), 20 participants inferred that the shopping icon is for on-line shopping and not a shopping list. Similarly only one participant inferred that time shown on the agenda icon (7:00) (Figure 1) referred to an appointment at 7:00 AM. These results confirm the compliance of 9 icons to the associated functions. They also provided us with valuable feedback to modify the icons of shopping list, home and office phone. Taking into account our observation that elderly people pay attention to details/indications of icons, adding a sheet symbol to the "shopping" icon improved the understandability of the icon.

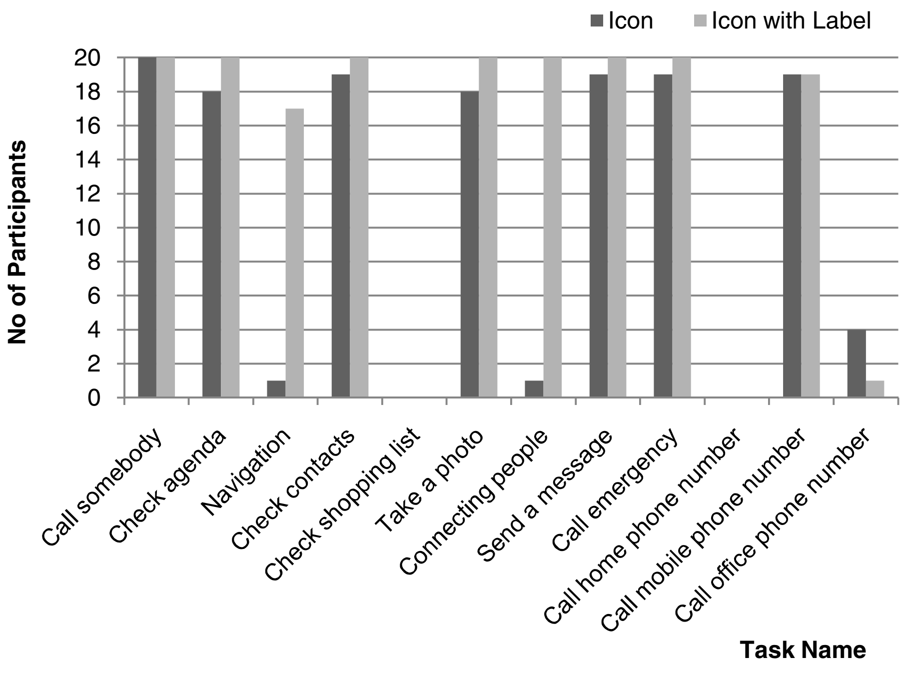


Fig. 3. Results of Icons Understandability Test

4.3 Use Test

Evaluation of PhonAge usage shows that the majority of participants (17/20) were able to easily execute the four proposed tasks (see section 2.1 Part3). Only 3 participants found these tasks difficult due to a visual impairment condition and their first time use of smartphones. During the first interaction with PhonAge, 17 participants used only the navigation arrows of the mUI to navigate through screens and icons selection (Figure 1-a). Only 3 participants used the touch scrolling, as they were already accustomed to it. After a demonstration time of the option (i.e. navigation arrows and touch scrolling), 17/20 participants preferred to keep both in the mUI. After this learning phase, participants found touch scrolling very intuitive as most of them said “we flip pages like in magazine, moving finger/hand form right to left.” 16/20 participants faced a difficulty to return to home screen (while navigation between screens). In the current version of PhonAge, returning to home screen is done always using the back button provided by the device. Since these participants were all novice Smartphone users, they were not aware of the presence of the button. One of the novice participants and the three who were already accustomed to touch screen devices had no problem returning to the home screen. After a learning phase of this option, the 20 participants were accustoming to the use.

The overall evaluation results shows that participants appreciated the clear color (e.g. green, yellow, blue) of the wallpaper and they appreciated the readability (16/20), visibility (20/20), pleasant aspects (18/20) and contrast (20/20) of the interface. Eighteen participants preferred the portrait display contrary to landscape display as they found it more classic.

The position of the icons and information on the main screen are also very much appreciated, for instance, all participants (20/20) liked the access to the spatiotemporal values and emergency button on all screens. Six participants also pointed, that one way to appropriate PhonAge is to personalize the label of the icons by adding the word “my” e.g. my contacts, my agenda, my shopping list rather than contacts, agenda, shopping.

4.4 PhonAge Compliance to Preferences and Needs of Elderly People

The use-test results confirmed the adequacy of PhonAge mobile interface design with elderly people perception. For most of the participants, the main purpose of using mobile phone is maintaining social relationships (calling and receiving calls, respectively 11/20 and 15/20) and having safety services (contact somebody in case of emergency, 20/20). The most useful PhonAge’s reported functions are: emergency (20/20), social service i.e. possibility to be connected with known and unknown people (17/20), and agenda to organize ADLs (15/20) (Figure 4 illustrates the results of PhonAge Service Usefulness).

In addition, participants found temporal information and its presentation very important; 16/20 reported that date is important while 14/20 reported time is important. However, 17/20 participants reported that weather information is not necessary to be available on all screens all the time. 15/20 participants reported that contact details is useful, they appreciated abstracted contact presentation (i.e. home, office and mobile contact information) (figure 5), and feel unnecessary to have detailed contact presentation (i.e. email, address, social contact).

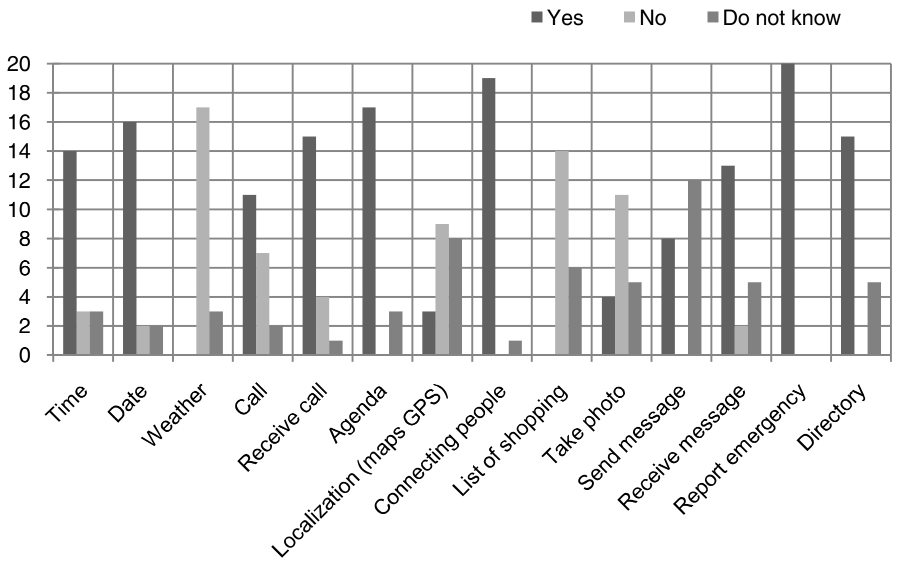


Fig. 4. PhonAge Service Usefulness

This feedback helped us to include the option of disabling detailed contact presentation in the configuration. Finally, only 3/20 participants (those already use smartphones) considered navigation service useful while other don't feel the need as their mobility is around their homes or they are accompanied with someone while going out. Only 4/20 participants find it interesting to take photos using smartphone camera, also all participants (20/20) appreciated the idea of displaying photos on the phone screen.

Fig. 5. PhonAge Contact Screen

5 Conclusion and Future Work

We presented in this paper our first evaluation of PhonAge, a smartphone based solution that comply with elderly people profiles. This evaluation confirmed that PhonAge is quick to learn and its design meets the expectation and perception of elderly people. Moreover, the feedback received from the evaluation helped to improve and modify the icons so that elderly can have a better understanding of the function they use. The evaluation also highlighted the concern of participants regarding the data entry for contacts, daily activities, shopping list, etc. In future, we are working to develop and evaluate PhonAge' services in the context of Age Friendly cities that would assist elderly to remain independent and maintain social interaction. We encourage research community to use PhonAge, and welcome an opportunity to work together. Our goal is to build a system that promotes successful aging by use of PhonAge as a host that enables successful deployment of useful assistive services.

References

1. Abdulrazak, B., Malik, Y., Arab, F.: PhonAge: Adapted Smartphones to Assist Elderly in Age Friendly Cities. In: 11th International Conference on Smart Homes and Health Telematics, Singapore (2013)
2. Abdulrazak, B., Roy, P., Gouin-Vallerand, C., Belala, Y., Giroux, S.: Micro context-awareness for autonomic pervasive computing. International Journal of Business Data Communications and Networking 7(2), 48–68 (2011)
3. Arab, F., Pigot, H., Rabardel, P., Folcher, V., Rigaud, A.S., Mokhtari, M.: Age, memory and time: practices and support. Journal of Association for Advancement of Modelling and Simulation Techniques in Enterprises 71(3), 136–149 (2011)
4. Batu Salman, Y., Kim, Y.H., Cheng, H.I.: Senior friendly icon design for the mobile phone. In: 6th International Conference on Digital Content, Multimedia Technology and its Applications (IDC), Seoul, Korea, pp. 103–108 (2010)
5. Fisk, A.D., Rogers, W.A., Charness, N., Czaja, S.J., Sharit, J.: Designing for older adults: Principles and creative human factors approaches, vol. 2. CRC (2009)
6. Helal, A., Mokhtari, M., Abdulrazak, B.: The engineering handbook of smart technology for aging, disability, and independence. Wiley (2008)
7. Oksman, V.: Young People and Seniors in Finnish 'Mobile Information Society'. Journal of Interactive Media in Education (JIME) 2 (2006)
8. Ziefle, M., Bay, S.: How to overcome disorientation in mobile phone menus: a comparison of two different types of navigation aids. Human-Computer Interaction 21(4), 393–433 (2006)

Experimentally Manipulating Positive User Experience Based on the Fulfilment of User Needs

Andreas Sonnleitner, Marvin Pawlowski, Timm Kässer, and Matthias Peissner

Fraunhofer Institute for Industrial Engineering, IAO
Nobelstr. 12, 70569 Stuttgart, Germany
{andreas.sonnleitner,marvin.pawlowski,timm.kaesser,
matthias.peissner}@iao.fraunhofer.de

Abstract. In this work, we prepared three variations of a prototype to experimentally manipulate parameters helping to improve User Experience (UX) of technological products. Based on a model considering the fulfilment of user needs (Fig. 1), two variations of a neutrally designed tool were developed to address two selected needs (*popularity, competition*) by slightly changing functionalities or design elements. The manipulation of UX is validated with real-time and retrospective subjective evaluation of UX, and objective data of user behaviour. Participants rated significantly higher positive UX and showed more active behaviour for designs of the prototype addressing the user needs c*ompetition* and *popularity* compared to the *neutral* design. These findings show the importance of considering elements of UX in the early development process of technological products.

Keywords: User Experience, UX, User Needs, User-Driven Innovation.

1 Introduction

User Experience (UX) is considered a key factor for the success of interactive products and systems. Looking at the definition of UX, it is “all aspects of the user’s experience when interacting with the product ... it includes all aspects of usability and desirability of a product ... from the user’s perspective” [1]. Next to this definition, various different interpretations exist due to the multidisciplinary nature of UX, ranging from a more psychological to a business perspective [2].

Our approach is based on the fulfilment of user needs during interaction with technological products derived from models describing general human needs [3,4]. The emphasis is placed on positive UX by meeting basic needs which is critical to the long-term success of technological products. Products that generated a lot of positive experiences will be used more frequently and will therefore be preferred [5]. In the context of work environments, UX can have a positive impact on skills and mental health and can also cause for motivated action, increased productivity and job satisfaction [6]. The developed model in Fig. 1 describes a holistic view of positive UX and will be taken as a basis for generating positive UX for interfaces. In addition to the final evaluation of a product regarding attitude and behaviour of users, also the

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 555–562, 2013.

experience itself and the cause for the use of the product are included in the detection of UX. Usability, directness and comfort are integrated as moderators.

The aim of this study is to specifically address user needs and to experimentally manipulate UX in a work environment, simulating both a more creative phase of idea generation, as well as a blunt execution of tasks. For this, an idea management tool is provided with differently shaped functionalities and design elements illustrating personal performance measures, idea evaluation and received achievements. In a first step, we decided to investigate two user needs with small overlap: *Competition*, i.e. to be better and outplay others, and *popularity*, i.e. to act dutiful and be liked by others. The influence of various functionalities and design elements on user behaviour as well as on real-time and retrospective subjective evaluation of UX was measured. The research questions investigated in this study are:

- Does the specific fulfilment of user needs lead to a higher UX?
- Can experimentally crafted functionalities and design elements satisfy the addressed user needs *competition* and *popularity*?
- Do experimentally crafted functionalities and design elements have an influence on user behaviour?

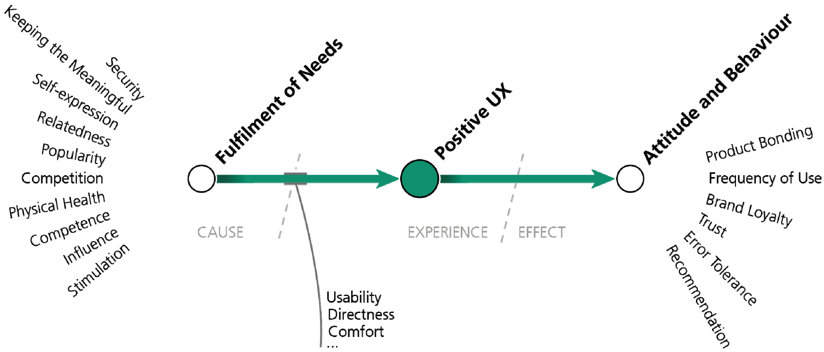


Fig. 1. Model of Positive User Experience

2 Methods

2.1 Participants

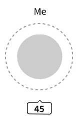
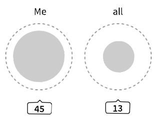
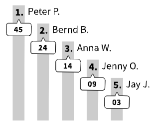
In the pilot study nine participants were involved to test first conceptual designs (18-74 years, mean: 35.9 years, 4 females). The sample of the main study consisted of 20 participants (18-70 years, mean: 38.5 years, 11 females, mean office working experience = 15.0 years). No subject participated in both studies.

Subjects were recruited from an in-house database in which volunteers for experiments are listed. Participation was voluntary, data were collected anonymously. Informed consent was obtained after the task had been explained. Participants were informed they had the option to end participation at any time without any type of penalty. After the study, they received 30€ for participation.

2.2 Pilot Study

As a result of wireframes and user centered paper prototyping, clickable prototypes of the fictional idea management tool were used to get information about elements most perceived by the participants and about user needs addressed particularly. With a combination of *Valence method* [7] and *Laddering technique* [8], emotions and needs were associated with functionalities and design elements. These results were used to improve the final design variations by focusing on salient elements with a high effect on user's perception (see Tab. 1).

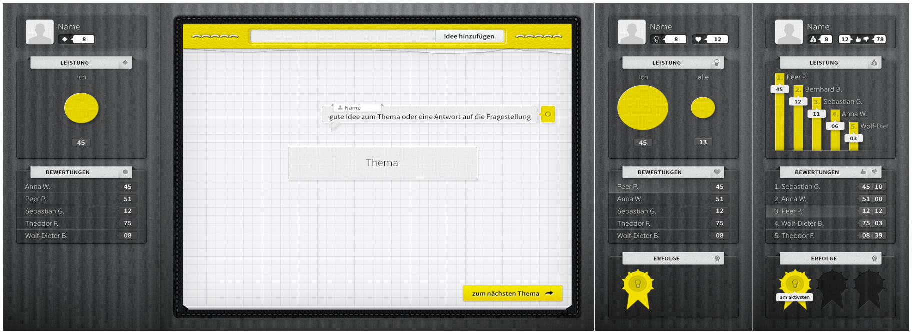
Table 1. Key differences for designs of the prototype (*neutral, popularity, competition*)

	Neutral (n)	Popularity (p)	Competition (c)
Idea evaluation concept	Isolated positive rating (likes)	Interactive positive rating (likes)	Interactive pos. and neg. rating (likes/dislikes)
Idea evaluation icon	 "circle"	 "heart"	 "thumbs"
Idea evaluation illustration	-	Alphabetic order	Ranking list
Performance concept	Own value	Own and team mean value	Own and each team member's values
Performance icon	 "diamond"	 "bulb"	 "bag of money"
Performance illustration	 "value based circle"	 "value based circles"	 "bar chart"

2.3 Main Study

Setting. To show coherences between design, functionality and UX, three prototypes of an idea management tool were developed based on Adobe Flash. Since the basic prototype is designed as a *neutral* one, the other two designs focus on fulfilling selected user needs (*popularity, competition*). Tab. 1 exemplarily shows some of the key differences of specific design and functionality characteristics. The aim of the study is to validate the assumption that slightly changing design and functionality of elements has an impact on the fulfilment of both user needs, and therefore has an impact on UX. The task is to brainstorm ideas to different topics in a team with four other simulated users that give input by a predefined script identical for all participants. Besides continuously brainstorming ideas to a total of nine topics (three per design), participants had to answer knowledge questions appearing on their screen.

On the main part on the right side of the screen, the current topic of the idea management tool is presented. Entered ideas are automatically arranged around the topic resembling the structure of a mind map. On the left side of the screen, personal information is presented including current performance measures, idea evaluation and received achievements. For entered ideas and correctly answered knowledge questions, the participants receive performance points. For idea evaluation, participants rate entered ideas by the team members positive and/or negative, depending on the functionality of the respective design (details, see Fig. 2 and Tab. 1).



2a 2b 2c 2d

Fig. 2. 2a and 2b exemplarily show a complete screen of the neutral design. On the main part on the right side of the screen the ideas were arranged resembling the structure of a mind map (**2b**). The key differences on the left side of the screen are drafted for the *neutral design* (**2a)** and the designs addressing *popularity* (**2c)** and *competition* (**2d**). See also Table 1.

UX Questionnaire. Based on the model of positive UX, a questionnaire was developed to measure the UX of a technological product that assesses the amount to which a need is fulfilled during interaction with a product. The *User Needs Questionnaire* (UNeeQ) consists of five items regarding overall positive and negative UX and 30 items regarding the ten user needs shown in Fig. 1. As a result, two overall scores for positive and negative UX, as well as ten scores for user needs are calculated.

Test Procedure. After filling out a small socio-demographic questionnaire, participants had to rank the three screens presented in Fig. 2 in terms of aesthetics without any additional information about the designs. For the main study, they started to pass through every design of the prototype in a randomly given order balanced over all subjects. At the beginning of each design variation, participants were given a standardized instruction how to use the respective prototype. After each design of the prototype, they were instructed to fill out the *UNeeQ*. After the main study, participants were again asked to rank the three screens presented in Fig. 2 in terms of aesthetics. To control the influence of individual competitiveness, participants had to fill out a German version of the Competitiveness Index (CI) at the end of the study [9].

Experimental Design. The study implied one independent variable, design of prototype (*neutral, popularity, competition*), and a covariate, user's competitiveness (*low, high*). The dependent variables were subjective ratings of UX (*UNeeQ*), subjective design-rankings (*preferences of designs*), and objective user behaviour (*frequency of likes/dislikes, number of ideas*).

A repeated-measures analysis of variance (ANOVA) was used for all within-subject comparisons to identify the effect of design of prototype and for between-subject comparisons to identify the effect of user's competitiveness on subjective ratings of UX and on objective user behaviour. Subjective design-rankings were investigated with Wilcoxon signed rank test. Only significant results are reported.

2.4 Hypotheses

- Participants report a higher *positive overall UX* in the *UNeeQ* for the designs addressing user needs *popularity* and *competition* compared to the *neutral* design.
- Participants report a higher fulfilment of the scales *popularity* and *competition* in the *UNeeQ* for the designs addressing user needs *popularity* respective *competition* compared to the *neutral* design.
- Participants show a more active behaviour for designs addressing user needs *popularity* and *competition* compared to the *neutral* design.

3 Results

3.1 Pilot Study

During Valence Method, participants marked elements regarding the idea evaluation the most (58 markers: 41 positive, 17 negative). Performance elements were marked 22 times (11 pos., 11 neg.), elements regarding the topic 17 times (9 pos., 8 neg.) and elements regarding the ideas 13 times (6 pos., 7 neg.).

The user needs addressed most with the clickable prototype were *popularity* (31), *competition* (21) and *security* (20). The other user needs mentioned were *competence* (7), *relatedness* (7), *stimulation* (5), *influence* (3), *idealism* (2) and *self-expression* (1).

3.2 Competitiveness

To control the effect of user's competitiveness, participants were divided into two groups based on the median of the total score in the CI of all 20 participants. ANOVA showed no effects for any of the dependent variables. Also, correlation between total score of CI and overall positive and negative UX of *UNeeQ* did not show significant correlations.

3.3 Subjective Ratings of UX

In Tab. 2, results of the ANOVAs are summarized. Overall positive and negative UX showed a significant difference for the variable design of prototype. Compared to the *neutral* design, overall positive UX was significantly higher and overall negative UX was reported significantly lower for the designs addressing user needs *popularity* and *competition*.

Single scales of *UNeeQ* showed a significant difference for user needs *popularity* and *competition*, as well as for the scale *influence*. All other scales of *UNeeQ* showed no significant difference for the variable *design of prototype*.

Table 2. Statistical results (ANOVA for repeated measures): Overall positive and negative UX and single scales of *UNeeQ* for the variable *design of prototype*

	Measure	Main effect		
		F(2,38)	p	η^2
Overall	Positive UX	10.415	<.001	.354
	Negative UX	7.534	.002	.284
User	Competition	11.637	<.001	.380
needs	Popularity	4.546	.017	.193
	Influence	4.469	.018	.190

3.4 Subjective Design-Rankings

At the end of the study, participants preferred the design addressing *competition* (mean rank = 1.40) in the subjective ranking over the *neutral* design (mean rank = 2.45) and the design addressing *popularity* (mean rank = 2.15).

Interestingly, the design addressing *competition* is ranked significantly better after the study compared to rankings before the study (Z = -2.762, p=.006). For the design addressing *popularity*, rankings were significantly worse after the study (Z = -2.233, p=.026). Rankings for the *neutral* design stayed equal.

3.5 Objective User Behaviour

To objectively compare user behaviour, frequencies of idea evaluation and number of ideas are compared. ANOVA showed a trend for the number of likes for the variable design of prototype ($F_{(2,38)}$=3.084, p=.057). Post-hoc analysis showed no significant differences between designs (mean number of likes: *neutral* = 6.05, *popularity* = 8.05, *competition* = 8.20). For entered ideas, participants showed no significant differences (mean number of ideas: *neutral* = 6.20, *popularity* = 6.70, *competition* = 6.45).

4 Discussion

In line with our hypotheses, overall positive UX, as well as single UNeeQ-scales for *competition* and *popularity* were significantly higher for designs addressing user needs *popularity* and *competition* compared to the *neutral* design. Also, *UNeeQ*-score for *influence* showed significant difference. A valid explanation is the obviously higher influence in the process of idea evaluation due to an interactive rating compared to an isolated rating in the *neutral* design.

Generally, user needs can only be separated to a certain degree. A small overlap already came up in the beginning of the development of our prototype when asking experts to attribute functionalities addressing special user needs. Nevertheless, all other scales showed no significant differences for the variable design of prototype. Therefore, the developed prototypes only addressed the expected user needs and showed a very good differentiation between single scales of user needs, which was an important requirement in our consideration.

Next to subjective ratings of UX, the two variations addressing user needs stated better spontaneous design-rankings compared to the *neutral* design. Interestingly, they preferred the design addressing user need *popularity* before the study, whereas after the study, they preferred the design addressing *competition* most. We assume that the slightly higher UX led participants to rank the design addressing *competition* over the design addressing *popularity* after interaction with the prototype.

In the pilot study, expected user needs *popularity* and *competition* were already most prominent in the clickable prototype. Besides, user need *security* was often mentioned when asking participants using Laddering Technique. It was mostly used for negative aspects of the usability of the clickable prototypes. Suggestions for improvement were considered and integrated for the prototype in the main study.

Regarding objective user behaviour, the design of the three prototypes had no significant influence on the brainstorming of ideas, but showed a tendency towards a more active idea evaluation for designs addressing selected user needs. We assume that the higher fulfilment of user needs, indicated by significantly higher *UNeeQ*-scores, led to this more active behaviour. Getting personal feedback due to interactive idea evaluation motivated participants to rate entered ideas by the team members.

4.1 Future Work

Future work aims at investigating all other user needs mentioned in the model of positive UX (Fig. 1). Variations of the described prototype will be developed to address the whole spectrum of user needs to define guidelines or specific parameters of elements to further improve UX for new products [10]. Another intended application of this prototype is to further investigate UX and emotions during interaction. Subjective surveys are usually conducted in retrospect by using questionnaires and interviews. The personal experience cannot be directly measured and is also shaped by personal attitudes and prejudices [11]. To measure UX during interaction with a technical product, neurophysiological parameters are promising to objectively indicate positive and negative emotions during experiencing [2]. They should help to explore the subconscious processes of implicit cognitions, and to draw conclusions about human behaviour [12]. Neurophysiological methods allow for a real-time measurement of brain activity during the current experience without significantly affecting the situation [13]. The approach chosen here is to identify and apply methods that provide insight into the subjective emotional experience in the interaction with technology and predict certain aspects of user behaviour.

References

1. ISO 9241-210: Ergonomics of human-system interaction - Part 210: Human-centred design for interactive systems (2010)
2. Roto, V., Law, E., Vermeeren, A., Hoonhout, J.: User Experience White Paper - Bringing clarity to the concept of user experience. In: Roto, V., Law, E., Vermeeren, A., Hoonhout, J. (eds.) Dagstuhl Seminar on User Experience - 2010, Dagstuhl (2011)
3. Sheldon, K.M., Elliot, A.J., Kim, Y., Kasser, T.: What's satisfying about satisfying events? Comparing candidate psychological needs. Journal of Personality and Social Psychology 80, S.325–S.339 (2001)
4. Reiss, S.: Multifaceted Nature of Intrinsic Motivation: The Theory of 16 Basic Desires. Review of General Psychology 8(3), S.179–S.193 (2004)
5. Kahneman, D.: Objective Happiness. In: Kahneman, D., Diener, E., Schwarz, N. (eds.) Well-being. The Foundations of Hedonic Psychology, pp. S.3–S.25. Russell Sage Foundation Press, New York (1999)
6. Harbich, S., Hassenzahl, M.: Using behavioral patterns to assess the interaction of users and product. Int. J. Hum.-Comput. Stud. 69(7-8), S.496–S.508 (2011)
7. Burmester, M., Mast, M., Jäger, K., Homans, H.: Valence Method for Formative Evaluation of User Experience. In: Halskov, K., Graves Petersen, M.G. (eds.) Proc. of the 8th ACM Conf. on DIS. Arhus, DIS 2010, pp. S.364–S.367. ACM, New York (2010)
8. Reynolds, T.J., Gutman, J.: Laddering Theory, Method, Analysis and Interpretation. Journal of Advertising Research 28(1), S.11–S.31 (1988)
9. Smither, R.D., Houston, J.M.: The Nature of Competitiveness: The Development and Validation of the Competitiveness Index. Educational and Psychological Measurement 52(2), S.407–S.418 (1992)
10. Sproll, S., Peissner, M., Sturm, C.: From product concept to user experience. Exploring UX potentials at early product stages. In: Proc. of the 6th NordiCHI: Extending Boundaries, pp. S.473–S.482. ACM, Reykjavik (2010)
11. Amelang, M., Schmidt-Atzert, L.: Psychologische Diagnostik und Intervention. Springer, Heidelberg (2006)
12. Bargh, J.A., Ferguson, M.J.: Beyond Behaviorism: On the Automaticity of Higher Mental Processes. Psychol. Bull. 126(6), S.925–S.945 (2000)
13. Spath, D., Peissner, M., Sproll, S.: Methods from neuroscience for measuring user experience in work environments. In: Karwowski, W. (ed.) Conf. Proceedings of the AHFE International, Miami, USA, July 17-20 (2010)

Floffy: Designing an Outdoor Robot for Children

Omar Mubin[1], Luke Vink[2], Pieter Oosterwijk[2],
Abdullah Al Mahmud[3], and Suleman Shahid[4]

[1] University of Western Sydney, Australia
[2] Eindhoven University of Technology, The Netherlands
[3] Department of Industrial Design, Delft University of Technology, The Netherlands
[4] Department of Communication and Information Sciences,
Tilburg University, The Netherlands
o.mubin@uws.edu.au, luke@lukevink.com,
p.j.m.oosterwijk@student.tue.nl, a.almahmud@tudelft.nl,
s.shahid@uvt.nl

Abstract. In our research we utilized the domain of entertainment robotics to educate children on the principles of environmental awareness by playful means outdoors. Our research revolved around the iterative design of Floffy: the environmental robot, which was essentially a playful toy robot that would respond positively to interaction that was beneficial for the environment and the child's own well being and negatively to interaction or behaviour that was detrimental to the surroundings. We conducted an explorative, informal evaluation of Floffy with two small groups of children and they rated their experience with it positively. Our results show that there is potential in utilizing entertainment robots to educate children on serious and critical issues such as saving our environment and being sustainable.

Keywords: Child-robot interaction, Arduino, Environment.

1 Introduction

In today's world dominated by industrial, mechanical and technological production, slowly sustainability and environmental awareness are gaining importance. Great emphasis has been laid forward on developing ventures that can educate people in protecting and preserving their environment. Educating our children on the importance of caring for the world we live in is also imperative, as awareness needs to start bottom up. The use of entertainment, games and robots to educate children is an exciting research area [1, 2]. LEGO Mindstorms is a well renowned learning tool, combining the aspects of play, games, robots and education [3].

The number of robots in our society is increasing rapidly and already in 2008 the sales of service robots outnumbered the sales of industrial robots [4]. Incorporating robots in behavioural revolutions therefore seems like an obvious choice. Gradually, users' interactions with robotic systems in their daily life will increase therefore there is great potential in utilizing those interactions to educate the users about various issues related to the betterment of the world. The notion of propagating environmental

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 563–570, 2013.

awareness with the help of robots is generally an untapped domain. An upcoming research stream is that of persuasive technology, where robots have been used to persuade users to save energy [5]. However what is measured is the power of verbal persuasion by robotic agents but there is no physical interaction involved with the robot. In addition there is no effect of the act of saving energy on the robot itself. Therefore our focus was on creating a robot that could educate its users on environmental protection, such that the robot was at the source of the environmental consciousness behaviour. The use of technology in general to instil environmental awareness in people is not new [6].

We wished to focus on child robot interaction for our research on creating a robot that would be environmentally friendly. Children of today will be the adults of tomorrow therefore any gradual behavioural revolutions in our society should also target children and not only adults [7]. There has been extensive research on educating children on the importance of preserving their environment [8]. One relevant prior work within the domain of environmental education using robots is that of [9], where they developed an ecologically friendly robot. However their target group was university students and the results of their evaluation are fairly limited, as only a questionnaire was used. The choice of age of our target group was another key decision. From prior literature [10], we deduced that at the age of 12 years old and more, children start to become more in touch with reality and develop a sense of reasoning and logic. In addition their cognitive thought processes become less egocentric and they start to think about their surroundings and the people in them. All in all, we felt that children of this age group would be willing to appreciate the principle of environmental awareness much better than younger children. This is further substantiated by the fact that children of the age of around 12 years would have probably gone through some environmental awareness programs at school or their local community. The intention of our research was not to replace such avenues but rather supplement them and provide a playful yet educational toy like product to children.

To summarize the design statement of our research was, "To design and evaluate an animated toy robot that in an organic form, represents appropriate animated behaviour to show a child around the age of 12 years old the importance of our natural environment." We realized that we would be unable to adapt a robot off the shelf (primarily because there is no such robot available that can effectively propagate environmental awareness to young children in a playful way). Our aim was to therefore indigenously create such a robot. In the following sections of the paper we first describe the iterative design process of Floffy and then the evaluation study that we carried out with children.

2 Design Process

Given that we were targeting a specific user group and that we had certain interaction trends in mind we outlined the following design decisions. These decisions were primarily informed from prior literature on entertainment robotics [11] (design decisions 1 and 2) and also from the brainstorming sessions we conducted with the research team (design decisions 3 and 4).

1. The robot should display certain animations (or behavior) to inform how it is feeling at a particular moment.
2. The robot should allow for personal, playful and tangible interaction.
3. The robot should physically represent a toy or a playful creature. However its representation need not be very literal, to allow for children to associate with easily, as children in the age group of 12 years old are likely to discover abstract signals and icons [10].
4. There must be elements of environmental consciousness built into the robot's interaction model. For e.g. aspects such as outdoors is better that indoors in a sustainable, environmental and general wellbeing sense (as evidenced in [12]), water is better than battery/water is renewable (as evidenced in [13]), etc. must relate back to the animations of the robot.

We followed an iterative process with the design of our robot. The first step of the design process was a series of several brainstorming sessions that we mentioned earlier. The following "physical attributes" of our robot were also brought forward as a result of our brainstorming sessions (along with the earlier mentioned design guidelines). The sessions were held within the research team (5 members) where one of the member took on the role of a facilitator and a second the role of the note taker. As part of the sessions we compared and contrasted existing entertainment products (toy robots) for children of our age group. Summarized below are the key attributes of our intended robot:

• The robot must move, i.e. leg/limb movement is an important design aspect to look into. A mobile robot would provide us with a richer array of interaction possibilities.
• The display of colors and lighting is also very important to utilize interaction possibilities.
• The robot should have an option to change shape.

The choice of prototyping platform was also a key design decision. We went ahead with the Arduino prototyping kit, which is a low-cost kit that comes with a microcontroller and various props that function as sensors and actuators. It is used fairly extensively in design research to create robotic prototypes [14]. We acknowledged that focusing on a solitary robot would not be an intelligent design decision. What was imperative was that we followed an iterative process, merging beneficial features from sub-prototypes. We now briefly present some of our intermediate designs (built using Arduino) (see Figure 1) that eventually led to the creation of Floffy.

Sub-prototype 1. The wagging tail. The Wagging Tail was an exploration of fluid movement to distinguish two opposing emotions (i.e. happy and sad) with two opposing movements, ultimately mimicking the shape dynamics of an animal tail.

Sub-prototype 2. The triangular robot. The triangular robot was again an expression of two opposing emotions. The triangle would be content in its default state (as seen in the figure). When suddenly approached by an unknown object in its proximity, the triangle would get scared and fall on its point.

Sub-prototype 3. Mr Googly. Mr Googly was an exploration of how the exterior of our robot could look like aesthetically and how it would feel in a material sense.

We specially focused on its big eyes, inspired by characters in children's TV shows such as in Sesame Street. The inside of the eyes contained RGB LED's. The choice of what colors in the LED's would depict what emotions was inherited from [15].

Sub-prototype 4. Walter Cannon. The Walter Cannon robot combined emotion and mechanics successfully using a proximity sensor to alternate its position and stance. If a user would draw his/her hand towards the robot it would quickly retreat.

Sub-prototype 5. Raffelsia. In order to explore techniques to portray the environment, we created the Raffelsia (flower like) robot. It consisted of an in-house built water sensor and an ultraviolet light sensor. The "petals" of the robot would open and shut depending on water intake and if it was outdoors or not.

Fig. 1. Sub prototypes (from L-R): The wagging tail, the triangular robot, mr googly, walter cannon and raffelsia

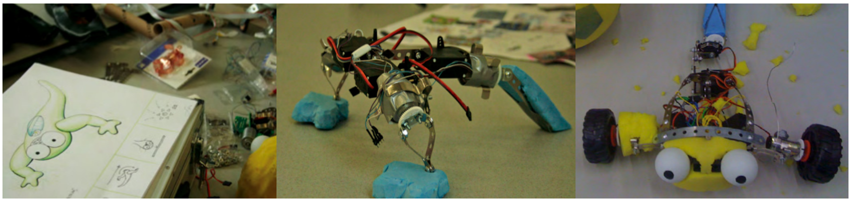


Fig. 2. Initial sketch of Floffy, the first structure of Floffy and redesign of Floffy (left to right)

3 Development of the Robot

Drawing on our experiences of creating the sub-prototypes we moved to the next stage of actually creating Floffy. We tried to incorporate facets of knowledge from the earlier design phase into our final design by utilizing the Questions, Options and Criteria Technique executed within the design team to decipher which features of the sub-prototypes would be the most critical for our final product. This led to for e.g. the water sensor mechanism of Raffelsia being retained for the final implementation. The first step was to sketch the prospective design of Floffy on paper (see Figure 2). We drew up several forms for the robot including 4, 5 and no legged creatures. Eventually, we decided on a robot with 3 limbs. This meant that whatever we worked out for the front two limbs, we could most likely make it move. The result was a gecko inspired creature.

The Personality of Floffy. Before creating the robot we agreed on an interaction scenario of Floffy based on its hypothetical personality: "Firstly, Floffy requires water to live. When Floffy is inside it responds with anger and sadness. It will jump and run away. When taken outdoors, its behavior changes and he begins to jump and play."

Building Floffy. Once we had an initial sketch and initial persona of Floffy in mind, the next step was to build the main frame/body of the robot, followed by adding sensors and finally the last stage of the building process was to focus on the physical attributes such as the outer skin and the padding to give it volume. We will now briefly describe each stage of the building process.

The frame of the robot was divided into modules, such as the two arms and the tail. While we built the individual modules we would predict the final shape of the robot with overlays to help in visualizing the outlook of the robot. Due to our success in the sub-prototype 5: Raffielsa, we wished to have arm movement in Floffy and we consequently arrived at the structure shown in Figure 3 (top). Due to a restriction on the number of ports on the Arduino board we used motors instead of servos.

We soon realized that due to a weight issue we could not achieve proper arm movements with the motors. They were unable to lift let alone move the robot. We also faced an issue with precise positioning of the continuous motor. Therefore our redesign involved the incorporation of wheels (see Figure 3 - bottom). The eyes of Floffy were adapted from sub-prototype 3: Mr Googly and the water sensor and solar panel from sub-prototype 5: Raffelsia. The movements of Floffy were based on our sub-prototype 4: Walter Cannon. The fur was also fairly complex to integrate because every part of the robot that moved had to be considered in conjunction with the placement of the water sensor and the solar panel.

The Behaviors of Floffy. Floffy, used a proximity sensor to determine in combination with a solar panel and water sensor, 9 different behavioral states. These were categorized into 3 groups determined by the current status of the robot. The 3 major states, 'near death', 'inside' and 'outside' were guided functionally but the existence of water (as food/energy) and then further broken down into 3 levels based on the proximity of the user. In the first major state; 'near death', Floffy would 'writher' in desperation (see Figure 5), if a user was in range (roughly 30 cm). As the person drew closer, Floffy would show a heightened and excitable desire for water by jumping swiftly. The second state, where Floffy was fed but inside, the robot would show negative behaviors, as an inclination to go outdoors. The closer the user got, the more Floffy would retreat and move away. In the final state; 'Outside', Floffy would show greater interest the more closer a user was by drawing itself towards the user. All the while, the robots eyes were changed in attempt to display these behavioral states. For example in a state of anger, Floffy would flash his eyes red, while joy or happiness showed the flashing of green eyes and desperation with orange.

4 Evaluation

Method. In order to evaluate our design of Floffy, we conducted a series of exploratory qualitative evaluation sessions with a group of children using the think aloud methodology. The think aloud protocol has shown to be effective with children [16] as compared to other standard evaluation methods such as questionnaires.

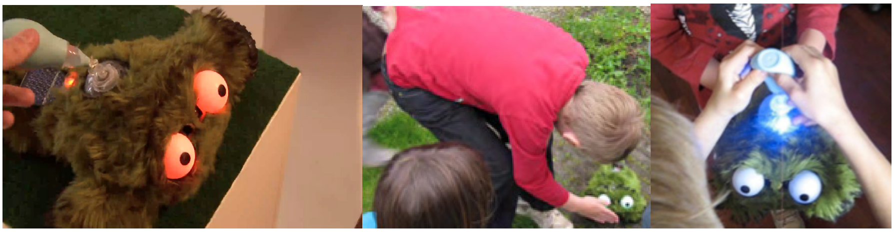


Fig. 3. Floffy showing desperation for water in the state of 'near death' (left most); Children interacting with Floffy

The evaluation was spread across two separate sessions, each involving 3 children (3M, 3F, average age: 12.2 years). Appropriate consent was gathered from their parents. The objective of the evaluation was to ascertain if the children could understand the rationale behind Floffy and interact and engage with it. The sessions started off by giving a small introduction of Floffy while indoors. The children were allowed free and open interactions (for 20 minutes) by bringing their hands close to Floffy (see Figure 6). They were also allowed to interact cooperatively with the robot and figure out how it all worked. At this juncture, the facilitators monitored what the feedback was from the children and also encouraged the children to elicit what they felt about each interaction step with the robot. The children were then given a water feeder without letting them know what they must do with it. The children would sooner or later discover that water would need to be dropped into Floffy (see Figure 6). As the facilitators we observed the reactions and thoughts of the children throughout the whole process. Floffy was then taken outside and the actions were repeated (free explorations for 20 minutes). We would like to share some interesting anecdotal notes from the evaluation.

Results. In the first session, one child attempted to pat Floffy from the front and due to Floffy's reaction with the proximity sensor, coupled with the robots lack of sunlight, it began to move fast and wildly (signifying an angry movement). The child became reluctant and took his hand away only to be later calmed by our explanation that Floffy is safe. In the second session, one of the girls exclaimed to us: *"He (Floffy) likes it better outside!"* When we questioned further as to why, the girls responded with, "Because his eyes are red inside and green outside". The children enjoyed playing with Floffy and 5 out of 6 children were able to correctly deduce what the purpose and intention of Floffy and its interactions was.

5 Discussion and Future Work

With our research we have investigated the impact robots can have on educating children in a fun and enjoyable way. We would like to summarize the key novel aspects of Floffy":

• Floffy is unique. No other robot on the market is “fed water” or “desires the sun”. This robot is designed to teach children to love the environment via its unique interaction mechanisms.
• Designed to be happier outside, children will prefer to play with Floffy outside at the same time learning to appreciate the environment the way Floffy does.
• Floffy convinces children into thinking it is powered by water, which is both inspiring for children and possible in the future. Floffy subtly suggests that the future of power is alternative energy.

One of the limitations of the work presented in the research so far is that to fully validate our claims we would need to carry out a large-scale evaluation with a substantial number of children. Therefore as part of our future work we would like to carry out extensive user evaluations with Floffy and children of 12 years old, which will help us in finalizing the granularity and specifications of the behavioral variations that we have implemented and also contemplate any redesign and structural modifications. We wish to conduct empirical evaluations of quantitative nature, enabling us to confirm the connection and link between the behavior of the Floffy and the interpretation of the child. We will also be able to establish what kind of actions and responses stimulate interaction between children and robots.

Acknowledgements. We thank the children and their parents for the cooperation**.**

References

1. Han, J., Jo, M., Park, S., Kim, S.: The educational use of home robots for children. In: ROMAN 2005, pp. 378–383. IEEE Press, New York (2005)
2. Fisch, S.M.: Making educational computer games educational. In: Interaction Design and Children (IDC 2005), pp. 56–61. ACM Press, New York (2005)
3. Mubin, O., Bartneck, C., Feijs, L., Hooft van Huysduynen, H., Hu, J., Muelver, J.: Improving Speech Recognition with the Robot Interaction Language. Disruptive Science and Technology 1(2), 79–88 (2012)
4. International Federation of Robotics.: World Robotics Survey (2008)
5. Midden, C., Ham, J.: Using negative and positive social feedback from a robotic agent to save energy. In: International Conference on Persuasive Technology, pp. 12–16. ACM Press (2009)
6. Morgil, I., Arda, S., Secken, N., Yavuz, S., Oskay, O.O.: The influence of computer-assisted education on environmental knowledge and environmental awareness. Chemistry Education: Research and Practice 5(2), 99–110 (2004)
7. O’Brien, K., Stoner, D.K.: Increasing environmental awareness through children’s literature. The Reading Teacher, 14–19 (1987)
8. Brynjegard, S.: School Gardens: Raising Environmental Awareness in Children. ERIC (2001)
9. Yamazaki, A., Masuda, R.: Development of an eco-communication robot for enhanced environmental education. In: International Symposium on System Integration, pp. 140–145. IEEE Press, New York (2010)
10. Markopoulos, P., Bekker, M.: Interaction design and children. Interacting with Computers 15(2), 141–149 (2003)

11. Graf, B., Oliver, B.: Entertainment Robotics: Examples, Key Technologies and Perspectives. Safety 6(7) (2002)
12. Coon, J.T., Boddy, K., Stein, K., Whear, R., Barton, J., Depledge, M.H.: Does participating in physical activity in outdoor natural environments have a greater effect on physical and mental wellbeing than physical activity indoors? A systematic review. Environ. Sci. Technol. 45(5), 1761–1772 (2011)
13. Frey, G.W., Linke, D.M.: Hydropower as a renewable and sustainable energy resource meeting global energy challenges in a reasonable way. Energy Policy 30(14), 1261–1265 (2002)
14. Alers, S., Hu, J.: AdMoVeo: A robotic platform for teaching creative programming to designers. In: Chang, M., Kuo, R., Kinshuk, Chen, G.-D., Hirose, M. (eds.) Edutainment 2009. LNCS, vol. 5670, pp. 410–421. Springer, Heidelberg (2009)
15. Russell, J.A.: Circumplex models of personality and emotions. American Psychological Association (1997)
16. Donker, A., Markopoulos, P.: A comparison of think-aloud, questionnaires and interviews for testing usability with children. People and Computers 16, 305–316 (2002)

Human-Spreadsheet Interaction

Andrea Kohlhase

Jacobs University Bremen, School of Engineering and Sciences, Germany
a.kohlhase@jacobs-university.de

Abstract. Spreadsheets have become very popular tools for analyzing and visualizing data from business and science. To better understand human-spreadsheet interaction, we explore readers' information models, but in contrast to most studies we focus on spreadsheet readers rather than spreadsheet authors. We conducted a study using the repertory grid technique and analyzed the result with the help of a Generalized Procrustes Analysis yielding a deeper understanding of human's information model of spreadsheets. Based on this we envision new human-spreadsheet interactions to increase the readibility and thus, usability of spreadsheets.

Keywords: Spreadsheets, repertory grid, information model of spreadsheets, human-spreadsheet interaction, information objects.

1 Introduction

The intuitive, flexible, and direct approach to computation in spreadsheets has led to widespread use and reuse. In particular, spreadsheets have become very popular to create, modify, and visualize numeric business and science data. In turn complexity and impact increased dramatically over the years. It has been estimated that each year tens of millions professionals and managers create hundreds of millions of spreadsheet programs [16]. This intensity yields not only more and more shared, complex spreadsheet programs, but also high-impact errors on the data level (up to 90% are affected according to [16]) and on the comprehension level (see [17]). The losses caused by formula errors and misinterpretation have even led to an international task force to battle them [4].

Previous Approaches. The standard research addressing such usability problems is based on Panko's influential report on error states and types in [16], that mainly considers the data level, e.g., computational errors based on faulty formulae.

Lewis and Olson gave a critical cognitive psychology account for the success of spreadsheets [12]. In particular, the barriers to programming are lowered, since the spreadsheet model can be used as visual programming language, enabling programming with low entry costs and early experience of success through effective displays and operations. Consequently, an abundance of research was directed towards end-user programming (EUP), i.e., "programming to achieve the result of a program primarily for personal […] use" [10, p. 4]. Unfortunately, from the point of view of

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 571–578, 2013.

EUP, a spreadsheet user is reduced to an end-user, that is simply any computer user who creates a spreadsheet program.

Spreadsheet Readers as Spreadsheet Users. But Nardi and Miller noticed another feature of spreadsheets in [15]: they are not "single-user applications". In particular, they are used in the work environment on the one hand as collaboration tool on the other as means of communication to exchange and combine domain knowledge and programming expertise. In [7] Hendry and Green highlight the fact that spreadsheet use is also a matter of understanding. They report their informants' missing comprehension (even of their own spreadsheet) and trace it to lacking comprehensibility support by spreadsheets, i.e., a specific usability issue for people reading spreadsheets. Such "readers" are, for instance, people that make use of existing templates by simply putting in new data, review data developments on different abstraction levels (like supervisors), assess data to base further decisions upon, want to understand their own spreadsheet program after a while, or look for reusable parts of a spreadsheet program, therefore browsing available ones. This means that an essential part of the spreadsheet user base is not yet addressed by current research. Many reports about bad decisions caused by misinterpretation and difficulties of spreadsheet comprehension confirm this observation (e.g. [2, 1, 19]).

Our Approach. Sometimes it is said that spreadsheets can be considered *data interfaces* that display and allow to play with data. Note though that data by itself is not interesting. In particular, Probst et al. suggested in [18] a knowledge management model. It differentiates information into four distinct traits. GLYPHS are just a set of characters without any structure, combined with a syntax they become DATA, additionally enriched by context they become INFORMATION, and finally, they turn into KNOWLEDGE if a semantic net or a global context is present. The question, in how far readers really consider spreadsheets merely as DATA interfaces, is the starting point for this paper. Moreover, we were interested in the evaluation criteria of readers with respect to a spreadsheet's information value. Based on such an analysis we are able to understand more deeply the information model of spreadsheets and can suggest innovative human-spreadsheet interactions.

Methodology: Repertory Grids and General Procrustes Analysis. Spreadsheet readers distinguish interface objects that carry information ("**information objects**"). To better understand what spreadsheet readers perceive as information units, what meaning they assign to these information objects, and how they discriminate between them, we conducted a study using the **Repertory Grid Interview (RGI) Technique** [9, 8]. RGI explores personal constructs, i.e., how persons perceive and understand the world around them. McKnight was the first to suggest RGI as a semi-empirical method for exploration of an information space [13]. By now RGI is a well-established method to explore users' personal constructs when interacting with software artifacts (see [21] for a list of examples). A crucial advantage over other methods is that a small sample size can be used.

A **repertory grid** is a grid consisting of "**elements**", i.e., the objects under consideration, and "**constructs**", i.e., pairs of antithetical properties that separate elements. The constructs serve as a bipolar dimension on which the elements are evaluated.

Elements as well as constructs can be elicited from the test persons themselves or can be provided by the interviewer. Comparison of multiple repertory grids is simplified if the individual ratings are given on a fixed set of elements or/and constructs, but a free elicitation explores the cognitive space. For our main RGI we decided to fix the set of elements, but to elicit individual constructs to better understand the information space.

We analyzed the repertory grid data obtained in the main study with "Idiogrid" and followed the analysis as described by Grice in [5]. In particular, we performed Gower's "**Generalized Procrustes Analysis (GPA)**".

2 The RGI Study[1]

In a first RGI we explored which information objects were discerned by spreadsheet readers in common spreadsheets. From this we extracted the most relevant ones to be included in the fixed set of elements for our main RGI study (see Table 1). Note that "diagrams" are missing, which may be due to the fact, that they were not part of our standard spreadsheet example. To broaden this set of elements, we added information objects which are not traditionally used in spreadsheets (see Table 2). In particular, we looked for such objects that contain spreadsheet-related information not usually available to spreadsheet readers. For this we made use of the semantic spreadsheet extension "**`SACHS`**" [11]. The union of both sets of information objects were used as the given, fixed set of elements, for which in this RGI study constructs were to be elicited by the interviewees.

Table 1. Common Information Objects of Spreadsheets

`Title`	A phrase describing the content of the spreadsheet
`Headers`	A (short) phrase supporting the interpretation of values of a regionally close range of cells (e.g. a column header)
`Legends`	A list of content properties and resp. layouts (as in a map legend)
`Values`	The content of a cell container
`Formulae`	A computational rule that yields a cell value
`(sx:)ColorCoding`	The use of color hinting at additional information
`Tables`	A possibly multidimensional homogenous structural layout of cells, that is perceived as an object of its own

Interview Procedure. In the study we presented each subject a simple, but complexly structured spreadsheet. Each element was explained by the interviewer and SACHS was introduced where necessary. Following traditional RGI, the interviewee was then handed three randomly selected element cards and asked to name one way in which two of the selected elements – considered as information objects – are similar or different from the other one. The label for the sameness was noted in the grid as left row

[1] A full description can be found at
`http://jpubs.jacobs-university.de/handle/579/2453`

header (the emergent pole), the label for the difference as right row header (the implicit pole) - yielding a construct. Then all elements were evaluated with respect to this construct with a binary rating scale: does this element rather belong to the implicit pole or the emergent pole?

Table 2. Extra Information Objects of Spreadsheets

`sx:LocalizedInfo`	A local look-up (data and text) of relevant information for cells on a by-cell-click basis
`sx:FunctionalBlock`	A local border indicating all cells functionally associated to the currently selected cell
`sx:DependencyGraph`	An overview graph (in a different window) of concepts showing on which the corresponding (selected) cell is ontologically dependent
`sx:RelationalArrows`	An arrow indicating a dependency relation between concepts in `sx:DependencyGraph`
`sx:ConceptNodes`	A node in `sx:DependencyGraph` representing a dependent subconcept, that additionally serves as a link to corresponding spreadsheet cells

Interview Summary. For our investigation we interviewed 14 people, of which 10 were male and 4 female. The age distribution was the following: 5 persons were under the age of 20, 6 between 20 and 30, 2 between 30 and 40, and 1 between 40 and 50. One subject had authored spreadsheets on a professional basis, 4 subjects were familiar with authoring simple spreadsheets, the other 9 only had occasional contact. All were explicitly asked to take up the role of a spreadsheet reader. Their background and education varied, but 3 were familiar with the `MS Excel` add-in SACHS before the interview. The rating scale was binary. In 1.5 to 3hrs sessions participants reported an average of 8.2 construct pairs (SD = 1.4) ranging between 5 and 11 pairs. A total of 115 constructs were elicited.

The first component of our GPA returned a rather high similarity (.68%). It was tested for statistical significance with the help of a randomization test based on 500 trials, which yielded an observed proportion $p \leq 0.00$ - verifying statistical significance. A subsequent standard **Principal Components Analysis (PCA)** produced $\{PC_{1...11}\}$. The first component explains ca 33.7%, the second 22.2% and the third 14.4% of the variance in the data. To approximate the meaning of the PCs, we looked at elicited similar – i.e., more salient (84%) – constructs and determined dimensions that can serve as common denominator constructs. As a result the PC_1 dimension is interpreted to range from "DATA Tool" to "KNOWLEDGE Tool" (interface perception), PC_2 aligns to "Represented DATA— Implicit KNOWLEDGE" (info perception), and PC_3 differentiates between spreadsheet use by authors or readers. The reliability of this qualitative analysis was ensured by following the procedure given in [8, 155ff.].

2.1 Towards an Information Model of Spreadsheets

Fig. 2 visualizes the element distribution according to the PCA, which we will discuss in the following. The x-axis corresponds to the interpretation of PC_1 as readers' perception of the interface, the y-axis to PC_2 as their perception of information.

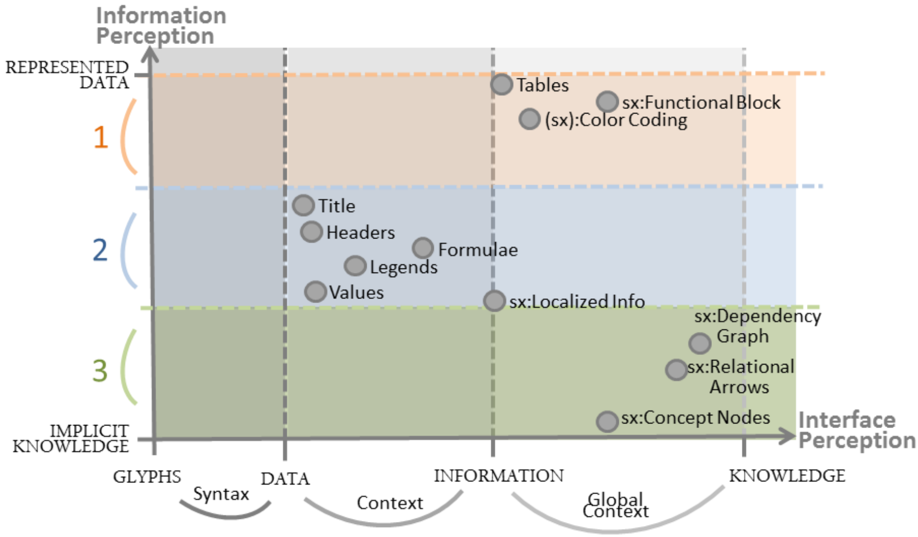


Fig. 1. Interpretation of Results

Perception Dimensions. Our interviewees perceived information objects differently when considered as tools (interface perception) or with respect to their information content (information perception). As tools, e.g., two sources can offer the same kind of information (`Tables` and `sx:Localized Info` provide INFORMATION), but with respect to their content they can be rated quite differently (`Tables` represents information as DATA whereas `sx:Localized Info` provides it as KNOWLEDGE).

Role Specificity. Our investigation showed certain information objects strongly associated with authors vs. readers. Interestingly, this indicates that readers only consider formulae (calculation), values (database data), tables (database views) and block arrangements (structural design) as creative options for the author. All the other information is intrinsically determined by the other elements. This suggests automation opportunities for spreadsheet applications of the future.

Information Environment. The information services offered by the SACHS extension and common spreadsheet applications do not overlap in the perception of readers. We can even acknowledge a progression between these element sets, where the common element set from Table 1 serves as a DATA-to-INFORMATION interface, whereas the SACHS set from Table 2 provides an INFORMATION-to-KNOWLEDGE interface (see Fig. 2).

Neighborhood of Information. The position of information sources inside or outside the frame of the application was explicitly observed by our interviewees. Elements were clustered according to the position of their respective point of reference.

Metaphoric Boundaries. In office suites the desktop metaphor prevails: Documents (like text files) are managed and can be accessed via players (like text processors). A player "plays" data, whereas a document "documents" data. For spreadsheets this very distinction by readers is rather surprising, as neither spreadsheet *programs*(!) are typical documents nor spreadsheet *applications*(!) typical players. Moreover, the desktop metaphor is rather limiting For spreadsheets, since the context-dependency of numbers is neglected. Finally, for readers the document/player metaphor is also restrictive as from their perspective the main purpose of spreadsheets consists in their communication, not in their documentation functionality.

3 Innovative Human-Spreadsheet Interactions

Based on the found dimensions we like to suggest new interactions to increase the usability of spreadsheets especially for readers:

1. The *perception of distinct dimensions of information objects* points to a frequently neglected media-theoretic topic that also concerns spreadsheets: Information objects are media and as such they do not only contain a message, they also are the message [14]. When using e.g. the information object `Tables`, then input data are perceived as DATA by readers. As DATA they need a context to become meaningful, but at the same time `Tables` as a structured, formal notation carries a 'truth' statement. Therefore, readers trust the information they get, even though the information object itself delivers no context to turn the DATA into INFORMATION. As a consequence authors should be compelled to create context, e.g. respective `Headers` or `Legends` if the spreadsheet is meant to be distributed, and readers should be required to understand the context before interpreting the DATA. The former is realized in many spreadsheet extensions/applications already, but the latter is not.
2. The perceived differentiation of spreadsheet users into *authors and readers* allows a much better fine-tuning of services. Even though the existence of both groups has been recognized, the interface design for players has not yet seriously taken this distinction into account. We will need more role-specific information services for readers. If readers, for instance, want to understand specific parts of a spreadsheet, these parts could be rendered separately on the fly as a spreadsheet view. This can reduce the cognitive overload when interpreting numbers in a big spreadsheet, particularly if information is scattered over multiple worksheets. Another reader specific service consists of a better navigation within spreadsheets, e.g. a semantically driven navigation as already prototypically presented with CogMap [6] or with SACHS' semantic navigation [11].
3. Our interviewees distinguished between *information environments* that turn DATA into INFORMATION and ones that turn INFORMATION into KNOWLEDGE. This induces the question how we can further enhance a data interface with "meta level" information objects. For instance, we could provide a reader access to the provenance of data or we could help the reader to assess information.
4. Following a *communication metaphor* for a reader, a communication mode of spreadsheets can be enabled, that provides on the one hand access to *document-specific* experts, background ontologies, or fora and on the other hand access to *topic-driven* discussions, domain knowledge e.g. in standard financial text books, help fora, or other domain services.

5. If we set the *document/player metaphor aside* then spreadsheets can also make use of already developed, open-standard, but non-spreadsheet-specific format guidelines: mathematical formatting of formulae.[2] For this, we can imagine a math editor and viewer, which takes input e.g. in LaTeX form – commonly used by mathematicians (which are typically non-programmers) for writing complex formulae –, converts it into MathML and renders it for reading in a browser window in standard mathematical notation. As our study indicated a neighborhood-of-information framing, we envision the window to be close to the cell for which such a formula is created.

Note that many of the envisioned interactions may be generalized to other office suite members to improve readability, that is, the communication aspect.

4 Conclusion

Interpretation and comprehension of spreadsheets constitute a rather neglected usability issue in research concerned with spreadsheets. In this paper, we presented a repertory grid study and subsequent General Procrustes Analysis that explore qualitative properties of information objects in spreadsheets from the point of view of spreadsheet readers. We discussed five framings of information sources in spreadsheets that readers perceived:

Perception Dimensions: Interface vs. information perception,
Role-Specificity: Information objects for authors vs. readers,
Information Environment: From DATA to INFORMATION with e.g. `MS Excel`, but from INFORMATION to KNOWLEDGE with e.g. `SACHS`,
Neighborhood of Information: Positioning of information sources inside or outside the frame of the application, and
Metaphoric Boundaries: A desktop vs. communication metaphor.

These dimensions of information from the readers' point of view represent relations between the set of information objects. Thus, they can serve as a first information model of spreadsheets and based on this we envisioned some innovative forms of human-spreadsheet interaction in the last section. As our study was only exploratory, we cannot conclude this information model to be complete. Nevertheless, it has become clear by this study that readers have their own interesting perspective on information offered in spreadsheets.

All in all, the information model of spreadsheets (by readers) presented in this paper is the entry door for a better, more complete understanding of human-spreadsheet interaction and a new source for according design inspirations.

Acknowledgements. This work has been funded by the German Research Council under grant KO-2484-10-1.

[2] We are fully aware that this might not be the best for a spreadsheet author, even though for complex formulae the typical spreadsheet formula language is not visual enough.

References

1. Caulkins, J.P., Morrison, E.L., Weidemann, T.: Spreadsheet errors and decision making: evidence from field interviews. JOEUC 19(3), 1–23 (2007)
2. Chambers, C., Scaffidi, C.: Struggling to excel: a field study of challenges faced by spreadsheet users. In: VLHCC 2010, Washington, DC, USA, pp. 187–194 (2010) ISBN 978-0-7695-4206-5
3. Dillmann, R., Beyerer, J., Hanebeck, U.D., Schultz, T. (eds.): KI 2010. LNCS (LNAI), vol. 6359. Springer, Heidelberg (2010)
4. EUSPRIG, European spreadsheet risks interest group (2010), `http://www.eusprig.org`
5. Grice, J.W.: Generalized procrustes analysis example with annotation (2007)
6. Hendry, D.G., Green, T.R.G.: CogMap: a visual description language for spreadsheets. J. Vis. Lang. Comput. 4(1), 35–54 (1993)
7. Hendry, D.G., Green, T.R.G.: Creating, comprehending and explaining spread-sheets: a cognitive interpretation of what discretionary users think of the spreadsheet model. Int. J. Hum.-Comput. Stud. 40(6), 1033–1065 (1994)
8. Jankowicz, D.: The easy guide to repertory grids. Wiley (2003) ISBN 0470854049
9. Kelly, G.: International handbook of personal construct technology, pp. 3–20 (2003)
10. Ko, A.J., Abraham, R., Beckwith, L., Blackwell, A., Burnett, M., Erwig, M., Scaffidi, C., Lawrance, J., Lieberman, H., Myers, B., Rosson, M.B., Rothermel, G., Shaw, M., Wiedenbeck, S.: The state of the art in end-user software engineering. ACM Comput. Surv. 43(3), 21:1–21:44 (2011) ISSN 0360-0300
11. Kohlhase, A.: Towards user assistance for documents via interactional semantic technology. In: [3], pp. 107–115 (2010)
12. Lewis, C., Olson, G.: Can principles of cognition lower the barriers to programming? In: Sheppard, S., Soloway, E. (eds.) Empirical Studies of Programmers, Norwood, NJ, USA, pp. 248–263 (1987)
13. McKnight, C.: The personal construction of information space. Journal of the American Society for Information Science 51(8), 730–733 (2000) ISSN 1097-4571
14. McLuhan, M.: Understanding media: the extensions of man. McGraw-Hill, New York (1964)
15. Nardi, B.A., Miller, J.R.: An ethnographic study of distributed problem solving in spreadsheet development, pp. 197–208 (1990)
16. Panko, R.R.: Spreadsheet errors: what we know. what we think we can do. In: [20] (2000)
17. Powell, S.G., Baker, K.R., Lawson, B.: A critical review of the literature on spreadsheet errors. Decision Support Systems 46(1), 128–138 (2008)
18. Probst, G., Raub, S., Romhardt, K.: Wissen managen, 4th edn. Gabler Verlag (1997)
19. Scaffidi, C., Shaw, M., Myers, B.A.: Estimating the numbers of end users and end user programmers, pp. 207–214 (2005)
20. Symp. of the european spreadsheet risks interest group (eusprig 2000)
21. Tan, F.B., Hunter, M.G.: The repertory grid technique: a method for the study of cognition in information systems. MIS Quarterly 26(1), 39–57 (2002)

Improving Students Learning Programming Skills with ProGames – Programming through Games System

Raquel Hijón-Neira[1,*], Ángel Velázquez-Iturbide[1], Celeste Pizarro-Romero[2], and Luís Carriço[3]

[1] Departamento de Lenguajes y Sistemas Informáticos I,
[2] Departamento de Estadística e Investigación Operativa, Universidad Rey Juan Carlos
[3] LaSIGE, University of Lisbon, Portugal
{raquel.hijon,angel.velazquez,celeste.pizarro}@urjc.es, lmc@di.fc.ul.pt

Abstract. We present a system for learning programming skills, ProGames, through a leveled set of visually-attractive and interactive programming exercises in Greenfoot, categorized by students likes to offer them solutions to problems they really enjoy or like most. The system has been evaluated during the course 2012-13 in 3 Computer Science Degrees and our results show very positive acceptance by the students.

Keywords: Programming teaching, Interactions Analysis, e-Learning, Moodle, Greenfoot, Visualization.

1 Introduction

The difficulty of learning abstract concepts, generally unknown to students, and the lack of systems proposing the unification between technical programming aspects and methods currently applied to foster student motivation, provoke in them lack of interest, failure and abandonment. That is why many concepts and methods of programming constructs are currently rethought. Professor Mark Guzdial [1, 2, 3] of Georgia Institute of Technology is considered a pioneer in teaching programming and sciences learning. He proposed a series of innovative activities aimed to enhancing the teaching-learning process in the engineering field. The main objective of ProGames, the system presented in this paper, is facilitating learning and foster motivation into the programming area. To this purpose, we propose a comprehensive set of programming games that are in fact exercises arranged into categories that students individually can select and complete according to their particular tastes. Therefore, there will be a wide range of solutions to problems in environments attractive to students where they will likely feel comfortably capable. Once a given student has been categorized, he will start his own path of learning proceeding progressively through the levels of the chosen category. Moving to the next level is possible only after successfully completing

* Corresponding author.

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 579–586, 2013.

all the tasks in the current one. The process can be repeated as many times as needed, until success is attained.

To provide visual interactive solutions to the programming exercises, the Greenfoot [4] development environment has been used because: i) its target audience is the educative community, and, ii) it includes tools for developing great visually-rich interactive applications. We have proposed 4 categories or scenarios for the problems. Each category consists of 7 levels of contents. Therefore, a set of exercises is proposed for each category and its 7 levels. All in all, a total of 192 different exercises are proposed. Students will download the set of exercises for each level, to be executed in Greenfoot Programming Environment. Thus, they can visually execute the interactive exercise solution, and understand it intuitively, providing a solution to a real case. Later on, students must study deeply the programming code, to understand it and become familiar with the different programming techniques used.

2 Existing Systems

One of the earliest pieces of software developed to teach programming is Logo [5], a high level language designed for teaching beginners to program. Another is Greenfoot [4], a restricted Java environment with custom high level libraries. In addition, much attention has lately focused on using a combination of visual programming with a 'storytelling' aspect, to make the experience of programming more immersive for children; namely Alice [6] and Scratch [7]. Letting children create their own stories in an open-ended way means that they can connect with their interests, which can be a key factor for motivation [8]. Using visual programming lets younger children become more easily involved in the process of programming both by removing syntax problems, and also through basic HCI mechanisms, such as recognition instead of recall.

A problem with simplified programming systems is that students often feel that they are not doing real programming, which can have a negative effect on both confidence and interest in programming. Related to this, is requiring students to make a commitment to learning the language and environment of the system. This is often somewhat wasteful since it can take a long time, and once they finish the course they will likely never use the system again – after all, these are not 'real' programming languages. There are a number of games, including LightBot [9], Robozzle [10], Carnage Heart [11] and VISPROCON [12] that make use of programming as the core game play mechanic. Although these games have a reduced programming fidelity compared to those mentioned above (e.g. Scratch and Alice), they do offer a lot of advantages. Games generally do not rely on teachers, which for the aforementioned reasons can be advantageous.

Project Euler is a series of challenging mathematical/computer programming problems that require more than just mathematical insights to solve. The trick is to craft a ladder of increasingly difficult levels, each one building on the last [13]. Project Euler is popular and engaging, in part because it is set up like a video game, with 340 fun, very carefully ordered problems.

With the right programming environment, students with even fairly rudimentary programming skills can produce software that is graphically impressive and supports complex user interactivity [14]. It is generally accepted in the literature that a programming assignment that involves building a real application is more rewarding than one that involves an isolated and unrealistic processing task [15].

The target audience of users of Greenfoot includes students from about 14 years of age, and scales up into introductory university education [4]. Users of Greenfoot may be complete programming novices within that age group, or they may have had prior programming experience in Scratch, Alice, or comparable systems. Concepts learned in Alice and Scratch transfer well into Greenfoot. The programming language used in Greenfoot is standard Java. The use of a textual language in general (as opposed to the Scratch/Alice model of drag-and-drop instruction blocks) is considerably harder for young learners to master. But we are using it at university level. Using Java in Greenfoot is considerably easier than using Java in a standard professional environment (be it a text editor/command line environment or a standard professional IDE).

3 The ProGames System Organization

To determine the student's preferences, we have created an ad-hoc website to offer a personal interest test based in the manual Kuder-C [16, 17]. This manual is applicable to teenagers and adults, so it perfectly adjusts the ProGames target audience. It offers an evaluation of interest of the subject on 10 different fields of interest with 168 questions on: Open Air, Mechanical, Calculus, Scientific, Persuasive, Artistic, Literary, Musical, Assistance and Administrative. The results obtained by a subject point out his likings or preferences for given type of activities. We have simplified it to the 4 more representative categories and only 20 questions. The chosen categories are: 1- Open Air: indicates preference for activities generally performed outdoors. 2- Artistic: indicates preference for creative works, generally dealing with (pleasing) works, such as drawing, coloring and attractive materials. 3- Assistance: indicates preference for activities that imply helping other people. 4- Calculus: indicates preference for numeric tasks and solving of mathematical problems.

The test would also give the affinity percentage to the other three categories. So the students, if desired, can complete his learning with the proposed exercises on the other categories, according with their preferences.

Each categories have seven levels of exercises that increase in difficulty: 1-basic programming constructs 2-structured instructions, 3- sub programming, 4-introduction to recursion, 5- arrays and algorithms with arrays, 6- files and records and 7- complex data types. He would firstly pick the category he has more affinity to, and start progressively with the adapted contents (lessons) for that category. In each of the seven lessons a set of exercises is proposed. The main idea is that the student downloads the exercises and executes them in the Greenfoot Programming Environment. Thus, he would visually check the exercise solution and, intuitively, understand it; so he has also a real use case for a given problem. Following to these, student must deeply study the exercise code, to understand it and learn the programming skills of each lesson.

The ProGames system has 192 exercises (49 for *open air*, 48 *Artistic*, 47 *Assistance* and 48 *Calculus*). These visually rich scenarios and the questions for their testing for learning programming have taken over 9 months of a full time Java programmer.

4 Example of Interactive Materials: Recursion with the Towers of Hanoi

The process for each exercise is that student would interact with the game, execute it several times, for different inputs and see what happens; once this phase is over, he would go to the code, and try to understand how the interactive game has been made. Next to that he would be asked a question about it, in order to check if he has understood it or not. In case of failure, he would have to repeat the process until the concepts become clear and answers are right.

Recursion is traditionally a very challenging concept for new programmers to master. Into the interactive Greenfoot world we are able to demonstrate recursion with the game "Towers of Hanoi". The game explains the movements required to transfer be made to move an entered number of discs from the first tower to the last, using as an intermediate tower the middle one. The player can move only a disc at a time, and a disc cannot be placed on top of a smaller one. Students can play the game varying the input data, which is the number of discs, just be clicking on the Act button (Figure 1).

Once the game has been played, student is requested to edit the Towers class to see the recursive function used (Figure 2). The variables origin, destination and aux correspond to the first, third and second tower respectively, and the n contains the entered number of discs for that execution.

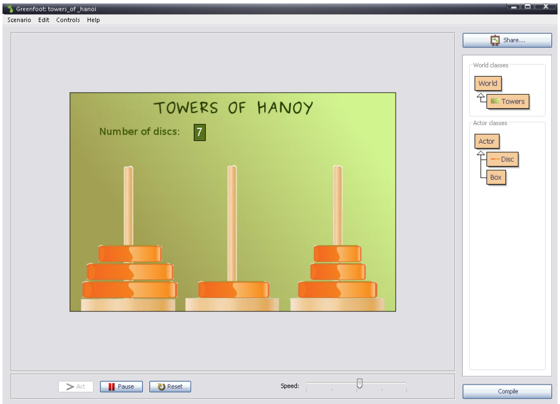


Fig. 1. Recursion with Greenfoot Towers of Hanoi Game

Students can now understand how a recursive function makes call to itself, and what are the parameters on every call. With some more games of this type in the selected category, the student will be ready to be examined by the virtual teacher in Merlin-Know, as explained in the next section, or on any other system.

```
private void moveDiscs(int n, int origin, int destination, int aux){
    if(n>=0){
        moveDiscs(n-1,origin,aux,destination);
        move(n,destination);
        moveDiscs(n-1,aux,destination,origin);
    }
}
```

Fig. 2. Source Code for the Towers of Hanoi Game

5 Acceptance by Students

The sample of students belonged to 3 Degrees of the Rey Juan Carlos University of Madrid: Computer Science Degree, Computer Science – Mathematics Double Degree, and Computer Science – Business Administration Double Degree. The three degrees had the same teachers, course materials and schedule. ProGames has been tested in the first semester (from September 20th to December 20th) of the academic year 2012-2013. The experiment included 44 CS1 (Computer Science) students of the 3 degrees of the course "Introduction to Programming", including: basic programming con-structs and structured instructions, sub-programming, introduction to recursion, ar-rays, files and records. We have combined the quantitative data provided by the ProGames system thanks to Merlin-Know and Merlin-Mo [18] and the subjective information provided by students through to a survey that students sent to teachers, using Google drive [19] technology, a survey on how students interacted with the system, to which degree they liked the various parts of it or how they reached it, is presented on this section.

The use of the ProGames was an extracurricular activity, only taken by 44 of the 117 students belonging to the 3 Degrees, who voluntarily wanted to; the use of the ProGames system was incentivized with the additional recognition of 1 ECTS credit when they passed with a grade of 7/10 all the tests in Merlin-Know (an ad-Hoc module for Moodle to test the progress in learning of students and that offers them feedback) of only one or two categories, 1.5 ECTS credits when have done the same with 3 categories and 2 ECTS credits when the 7/10 grade was obtained for all the tests of the 4 categories.

The data analysis has been done with Many Eyes [20] that allows interactive analysis. Figure 3 presents the analysis of categories, on the x-axis the seven lessons for each category (from 1 to 7), on the y-axis, the 4 categories (from bottom to top: *Assistance, Calculus, Artistic* and *Open Air*), the bubble size indicates the number of students that chosen it, and color indicates the acceptance grade (legend is on the top right). On the categories analysis, the first three were clearly more picked than the last one, *Open Air* which is the last one that student completed, about ¼ less students completed it (44 to 29 approximately). It clearly seems that the last of students 'choice was Open Air, which can be positive for future computer science professional's which work will mostly be done indoors. If exercises are proposed for students on these 3 Computer Science Degrees, on the first three categories, they will probably have better acceptance. Taking into account the lesson analysis for categories, it clearly seems that from lesson 4 onwards, there are a higher students' percentage 'not liking much' the lessons; we think it can be because the difficulty grade increases, since these lessons are: recursion, arrays and algorithms with arrays, files and registers and advanced data types. Actually, first three lessons are more liked by students, particularly lesson 3 in the calculus category, which has had absolutely no dislikes. On the contrary, almost ¼ of students did not like lesson 7 in the Artistic category, whose exercises should be improved on future ProGames versions. All in all, lesson 3 in all categories has been exceptionally well received; it is the one corresponding to subprogramming, which is a very important concept taught on an introductory programming university course.

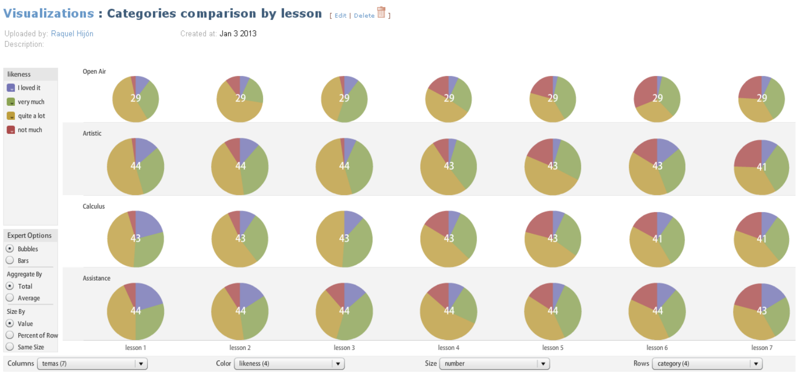


Fig. 3. Scatterplot of categories comparison by lesson, size and likeness

Figure 4 shows students preferences selected on the Kuder-C test [16, 17] of vocational preferences taken by students for their categorization at the beginning of the experiment.

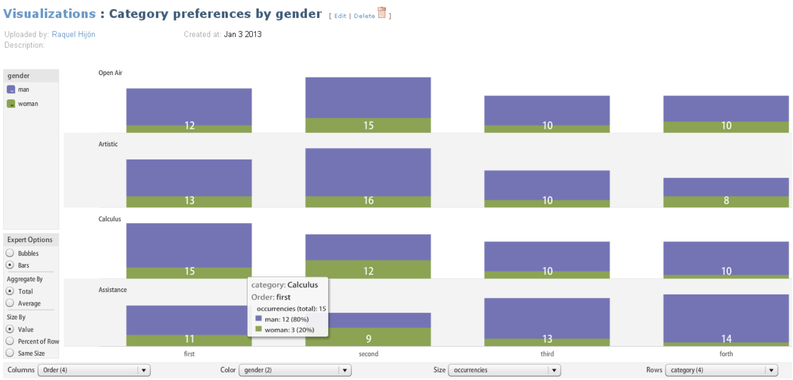


Fig. 4. Scatterplot of Categories chosen by genders

On the x-axis is shown the order of categories' preference they had marked on the initial test based on it, and therefore followed into the ProGames system. On the y-axis the bar size indicates the number of occurrences, and the color differentiates between men and women. The category chose on first place has been calculus, for both men and women. Therefore, students on Computer Science Degrees show preference for these types of activities, so if we offer them this type of problems, they will accept them better. It is very closely followed by the Artistic an Assistance categories. Considering the gender analysis, the percentage of women liking the Assistance category on the second place is considerably higher than in any other category. Anyway, their first choice is also calculus, as it is also for men

For the interfaces definition task, it is very important to determine how students access to the ProGames system. Thy used PCs (40%), laptops (43%), notebooks (9%), smartphones (5%) and tablets (4%). The reason why the smartphones' percentage is so low is because the Greenfoot API needs to be installed on a computer in order to be able to execute the exercises.

When at the end of the evaluation students were asked, 'in general have you liked working with the ProGames system?" they answered that quite much (48%), very much (25%), a little bit (20%) and not at all (8%).

6 Conclusions and Future Work

It was a challenge to think on a system for learning programming that fit the students' personal needs and liking, aimed to improving their learning process. Furthermore, it was our goal to offer the 192 leveled sample exercises in an environment interactive, visually attractive and through programming games. It was very important being able to offer the ProGames system to students in a LMS that nicely supported it. Finally, it was important for us being able to test the progress of students through their learning into the system, to analyze their interactions, and also show them their personal progress, performance and comparison with their classmates. Those objectives have been fully satisfied. Firstly, the categorization has been possible thanks to the development of a web system based on Kuder C test of personal interests. Secondly, the interactive and very attractive to students' game-based programming environment for the development of the 192 programming exercises has been Greenfoot. Thirdly, the LMS system proposed hosts the system the collection, Moodle, and the new module developed for it, Merlin-Know, with its virtual teacher that guides and motivates students learning and show them their progress. As a whole, we feel very satisfied with the solution proposed.

ProGames system has been tested on 3 Computer Science Degrees of Escuela Técnica Superior de Ingeniería Informática of Universidad Rey Juan Carlos during the academic year 2012-2013. The students' interaction analysis and their subjective opinion about the system have been very satisfactory. Almost all exercises on almost all categories were very much enjoyed, successfully accomplished and accepted, which has arisen a very positive opinion on them.

More solid studies are carried out to compare possible differences in increase on knowledge between ProGames and the traditional learning method.

References

1. Rich, L., Perry, H., Guzdial, M.: A CS1 Course Designed to Address Interests of Women. In: Technical Symposium on Computer Science Education, Proceedings of the 35th SIGCSE Technical Symposium on Computer Science Education, Norfolk, Virginia, USA, pp. 190–194 (2005)
2. Tew, Fowler, Guzdial: Tracking an innovation in introductory CS education from a research university to a Two-Year College. In: Proceedings of the 36th SIGCSE Technical Symposium on Computer Science Education, St. Louis, Missouri, USA, pp. 416–420 (2005)
3. Ericson, Guzdial, Biggers: A Model for Improving Secondary CS Education. In: Proceedings of the 36th SIGCSE Technical Symposium on Computer Science Education, St. Louis (2005)
4. Kölling, M.: The greenfoot programming environment. ACM Trans. Comput. Educ. 10(4), Article 14, 21 pages (2010)
5. Logo Foundation, MIT University, `http://el.media.mit.edu/logofoundation/index.html` (accessed March 25, 2013)
6. Alice.org, Carnegie Mellon University, `http://www.alice.org/` (accessed March 25, 2013)
7. Scratch Home, MIT University, `http://scratch.mit.edu/` (accessed March 25, 2013)
8. Koster, R.: A Theory of Fun for Game Design. Paraglyph Press (2004)
9. light-Bot, Armor Games, `http://www.kongregate.com/games/Coolio_Niato/light-bot` (accessed March 25, 2013)
10. RoboZZle online puzzle game, `http://robozzle.com/` (accessed March 25, 2013)
11. Carnage Heart, `http://en.wikipedia.org/wiki/Carnage_Heart` (accessed March 25, 2013)
12. Bromwich, K., Masoodian, M., Rogers, B.: Crossing the game threshold: A system for teaching basic programming constructs. In: 13th International Conference of the NZ Chapter of the ACM's Special Interest Group on Human-Computer Interaction, CHINZ 2012, Dunedin, New Zealand, July 2-3, pp. 56–63 (2012)
13. Somers, J.: The Atlantic (2013), `http://m.theatlantic.com/technology/print/2011/06/how-i-failed-failed-and-finally-succeeded-at-learning-howto-code/239855/` (accessed March 25, 2013)
14. Haden, P.: The incredible rainbow spitting chicken: Teaching traditional programming skills through games programming. In: Proceedings of the 8th Australian Conference on Computing Education, pp. 81–89. Australian Computer Society, Darlinghurst (2006)
15. Huang, T.: Strategy game programming projects. The Journal of Computing in Small Colleges 16(4), 205–213 (2001)
16. Kuder, G.F.: Professional preferences, Form C, Vocational, Tea Ediciones (1986)
17. Kuder, G.F.: The Stability of Preference Items. Journal of Social Psychology 10, 41–50 (1939)
18. Hijón-Neira, R., Velázquez-Iturbide, J.Á.: Merlin-Mo, an interactions analysis system for Moodle. In: Proceedings of the 16th Annual Joint Conference on Innovation and Technology in Computer Science Education (ITICSE 2011), p. 340. ACM (June 2011)
19. Google Drive, `https://drive.google.com` (accessed March 25, 2013)
20. Many Eyes IBM Research and the IBM Cognos software group, `http://www-958.ibm.com/software/analytics/manyeyes/` (accessed March 25, 2013)

Long-Term Experiences with an Iterative Design of a QR-Code-Based Payment System for Beverages

Max-Emanuel Maurer, Alexander De Luca, Alina Hang, Doris Hausen, Fabian Hennecke, Sebastian Loehmann, Henri Palleis, Hendrik Richter, Simon Stusak, Aurélien Tabard, Sarah Tausch, Emanuel von Zezschwitz, Franziska Schwamb, Heinrich Hussmann, and Andreas Butz

University of Munich (LMU), Institute for Informatics, Munich, Germany
firstname.lastname@ifi.lmu.de

Abstract. We report on the design and long-term use of a digital tracking system for the consumption and payment tracing of beverages, called "Barkeeper". It is based on tags wearing QR-codes and its design was not primarily guided by efficiency, but rather everyday use during the last three years in our lab. In this trusted environment, we collected extensive usage data, making this a serious long-term field deployment of UbiComp technology. We present the system, its iterative design evolution, the users' views on it and insights gained by daily usage. We argue that QR-code interaction, when implemented in a very pragmatic way, can be not only a cheap but also a very powerful interaction technique. Based on our experience we propose a set of general rules, which make QR-code-based interaction practical and often superior to other interaction techniques.

Keywords: visual markers, QR-codes, ubicomp, payment system.

1 Introduction

Caffeine is an essential part of every (good) computer science lab. A number of studies proved the positive effect of caffeine on performance [2]. As in many places, members of our group have to pay for their office consumptions introducing the need to track the expenses of each coworker. Motivated by practical issues with existing systems, we developed and tested a number of (technology-mediated) solutions over the past years to keep track of beverage consumption and share the cost.

Our starting point was a shared paper tally list next to the coffee machine where consumers marked their consumption. The tally list proved to be error prone and generated an increased workload as the group grew larger. An open money jar was incompatible with the public nature of our lab, with many students and unknown people passing by. Through an iterative process driven by our own needs and everyday use, we developed a digital payment system: the Barkeeper. It has been used daily for over three years and is well accepted by all members of the group.

When designing the system, we originally considered touch screens, RFID [4, 5], NFC [6] or even sensing cups [1], but we finally settled for a visual marker,

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 587–594, 2013.

i.e. QR-code, scanner for interacting with the system. QR-codes are (i) cheap, (ii) robust (in particular if coated with transparent plastic), (iii) suited to make the system highly scalable (e.g., when adding people), and (iv) easy to use. In comparison to other systems, the Barkeeper does not require users to actively take pictures, which is a common approach. Instead, our users place their tags under a fixed camera. To pay for an item, users scan two tags consecutively in any order: one representing the user identity (see figure 1), the other representing the product.

We present the results from three years of everyday interaction with this system in our lab. To our knowledge, there is no work in which visual markers were evaluated over such a long period of time in a real-world deployment. We discuss the reasons why the system is so well accepted and in particular the advantages of using QR-codes in the way our system does. We also report insights from different user perspectives. This includes highly personalized strategies of the payment system (e.g., different ways to scan and store visual markers) and effects on social interaction within the lab. We will also draw conclusions that are generalizable beyond the scope of our coffee kitchen. For instance, we found that spatial proximity highly influences the use or non-use of such a system, even if the difference is only a few meters.

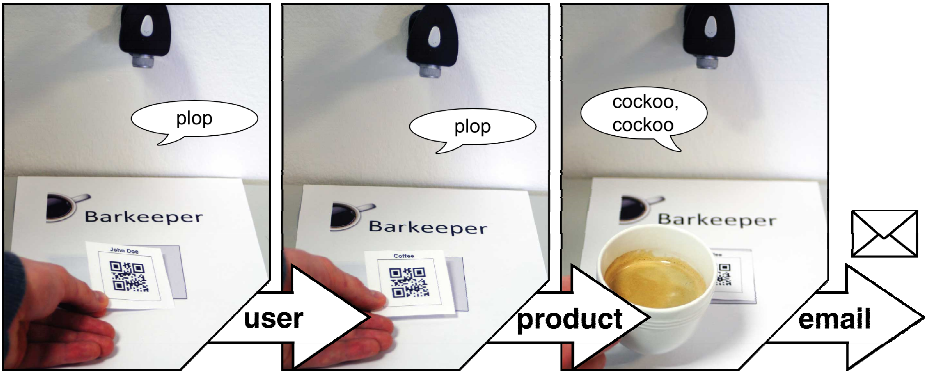


Fig. 1. Interactions performed with the Barkeeper when buying a drink. After consecutively scanning a user and a product tag (or vice versa),sound feedback indicates the successful transaction with additional feedback sent via email.

2 The "Barkeeper" System[1]

The driving idea behind the system was to improve compliance by having users pay as they take a drink, while decreasing the accounting and management load. The Barkeeper is composed of a webcam-based scanning application on a thin client talking to a webserver. The application waits for a user to scan two tags consecutively (one person and one product tag). Once both types of tags have been scanned, the transaction is recorded. A simple auditory feedback indicates whether a scanning action was

[1] You can find a video presentation of the system online at: `http://youtu.be/NqmEqC24HCw`

successful or not (e.g., system resets due to a time constraint or submission failure). As additional feedback, users receive an email after every purchase. We decided to avoid visual feedback through a monitor since the system is installed in a kitchen (prone to dirt and moisture problems). Charging user accounts is done through an administrative web interface.

The everyday usage of the Barkeeper led to many suggestions for improvements. While exploring the design of a payment system, we found that using the system ourselves also gave us a better sense of what mattered and pushed us to consider how the Barkeeper integrated in the lab's life [3], as we describe below by outlining the evolution of the Barkeeper.

The first change happened only after a few days: to save time we introduced so-called "combi-tags", a single tag combining a user id and a product id. This enables users to print combi-tags for their favorite beverages. After a few months we introduced "group tags": a special kind of person tag serving as an alias for a project or group of people whose consumption is usually covered by project funds. We also introduced a guest tag whose drinks are paid for by the entire lab.

Further enhancements include a transaction feature to transfer money between accounts. For example, if two colleagues went to lunch together and one of them paid for both, the money can afterwards be paid back via the Barkeeper. We also adjusted the auditory feedback and replaced the original horn-like sounds by more natural ones. Successful scanning is now indicated by the call of a cuckoo and an error by the cark of a crow. These sounds are distinct enough to function as signals for the actual user and at the same time unobtrusive enough to not distract colleagues[2]. Besides an online frontend to check the balance, print combi-tags, transfer money, and change several settings we also developed an Android app to purchase drinks, make financial transactions and check our balance.

3 Living with the Barkeeper

Today, our group consists of 18 members and is spread out over three of the six floors in the building. Thus, depending on the location of their office, people have to cross different distances between zero and four floors to get a beverage. To manage this particularity, our heterogeneous group of users developed several unique strategies. Using an online survey and semi-structured interviews we gained insights into these individual strategies.

3.1 User Perspective: Daily Life with QR-Codes

A survey with 21 current and former users (six female; 31 years avg.) yielded results about usage patterns. 62% of the participants stated that they used the system daily, 77% of these even multiple times a day. The preferred interface for buying drinks is

[2] As a subtle side effect, one user reported being now irritated when hearing a crow in nature, wondering what went wrong.

the tag-based payment, which is used frequently by the majority (81%). Nearly all (95%) hardly ever use the web interface and only two (10%) use the mobile app. Several styles of tag usage for payment have emerged: most users (38%) search for both – person tag and product tag – first and then place them under the camera consecutively. Others (19%) find one tag, place it under the camera and use the scanning time for finding the second tag. One user notes: "I always kind of conceived it as a game: do I succeed in finding the second tag before the purchase would fail due to a timeout?"

Seven participants (33%) speed up the process using at least one combi-tag. Five participants (24%) reported other methods such as attaching a small flag to their name tag or changing the location of their personal tag (e.g., sticking it onto the kitchen cabinet or storing it at one's office desk). Furthermore, people print tags on colored paper to make them more easily distinguishable. People even started to optimize their processes in the kitchen according to the Barkeeper paying process (e.g., starting coffee brewing before scanning the tags). The Barkeeper can even act as a decision helper: "Sometimes, the first drink tag I see influences my choice of product."

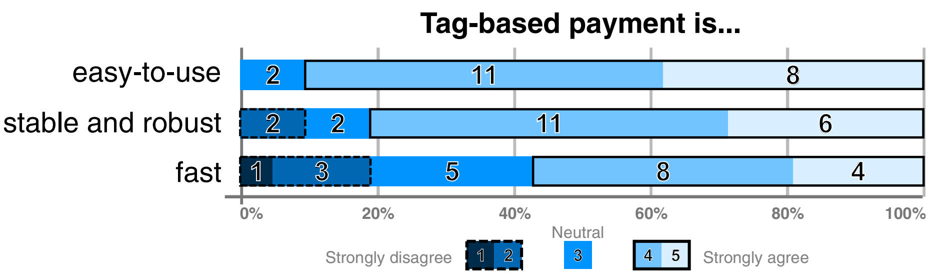


Fig. 2. Rating of the tag-based payment

In general, participants are satisfied with the Barkeeper. The tag-based payment is rated as easy-to-use (mode=4 on a five-point Likert scale ranging from 1=strongly disagree to 5=strongly agree), stable and robust (mode=4) as well as fast (mode=4) (see figure 2) with some people speeding up interaction by using combi-tags. Besides the initial barrier of printing these tags manually, there seem to be different reasons for that. Participants apparently differentiate between fetching a quick drink and taking a larger break from work. Although we did not find any significant differences in our questionnaire results, this also seems to correlate with the three different floors people are working on. People on the lower floors (without coffee kitchen) possess less combi-tags as they have to come up three or four floors to get their drinks and mostly seem to do this for larger breaks. Colleagues working on the upper floor (where the kitchen and the Barkeeper are located) do their purchases in between their work and more in a hurry, using combi-tags more often.

In terms of efficiency, the Barkeeper introduced the biggest improvement for the administrative staff in our lab. The old tally list required the secretary to take it down, calculate all unpaid amounts and put a new list up every week. The only task now is to accept the money from users, put it in the physical cashbox and add the paid amount to the respective Barkeeper account using the administrative web interface.

As one secretary stated, the most complex thing for her while working with the Barkeeper is “to put the cash into the cashbox”.

3.2 Security, Privacy, Trust and Social Implications

Obviously, the Barkeeper is not secure against misuse. A photocopy of a tag or the knowledge of its contents is enough to fake tags for malicious purchases. However, security can be easily increased by keeping the tag with oneself. In comparison with a tally list, the system offers much higher security and privacy. Wrong tally marks have been undetectable and occasional incorrect purchases inevitable. Due to the purchase confirmation email as well as an overview with all purchases in the web interface and in the mobile app, wrong bookings are easily detectable. From a social perspective, the daily use of the system influences group behavior. Being used for interpersonal money transfer, the Barkeeper has established a new verb in communication between lab members. “You can just barkeeper that to me” has become a phrase commonly used. The data of the Barkeeper is also used as a means of social pressure to support positive payment behavior. Every 30 days, the lab employee with the highest debt is announced in an email and is forced to buy milk for the coffee for the next 30 days. This caused the cash box to change from being constantly empty to being well filled.

3.3 Benchmarking

To assess the usability of the different booking methods (tags, web and paper), we measured average interaction times with each of them.

Methodology. We measured the average interaction times with single tags, the web frontend and the tally list of twelve users. The Android app was excluded due to underutilization. Measurements were done with a stopwatch during daily routine to capture influences like the changing order of tags in the collecting box. Hence the measured times include searching the right tags as well as successful scanning. The interaction with the tally list includes the search for the pen as well as for the right row and column. Possible personal strategies were recorded.

Results. Average interaction time with the tally list was fastest (m=7s; sd=2s), followed by the web frontend (m=15s; sd=3s). The interaction time with the single tags was slowest (m=22s; sd=7s). Only four participants used combi-tags, although using them was faster than the tally list (m=6s; sd=1s).

The analysis of the Barkeeper database is in line with the qualitative results. Until December 2012, 11,941 QR-based transactions were performed, while 1,188 web-based transactions (10%) and only 76 transactions using the mobile app (0.6%) were logged. Since its introduction, the Barkeeper was used on 96% of all work days (excluding public holidays) with a total average of 20 transactions per day. During this time, a total of 14,878 € have been transferred.

4 How to Make QR-Codes Work

The benchmark results show that the introduction of the system produced measurable disadvantages for its users in terms of time. However, users accepted the transition, continue to use the system daily and even stick to the slower procedure using two single tags. Hence we have experienced a difference between subjective experience of time and objective interaction time. In the following paragraphs, we will point out a few general findings, which can be applied beyond our special use case.

4.1 Performance Is Not Always the #1 Priority

In HCI research, interaction speed is one of the most used dependent variables for measuring system quality. However, in some scenarios like our system, this might not be as important as a subjective experience of time. Being offered diverse modalities (tags, app, Internet) and various ways to speed up interaction (e.g. combi-tags) only a fraction of the users adopted those possibilities and preferred using the tags. This is probably due to the integration of the interaction into the process of taking a break and due to the generally playful experience. We argue that systems that are able to "hide" time costly interactions within other processes or through playfulness can still be well accepted even while being objectively slower.

4.2 A Static Camera Makes QR-Codes More Usable

One of the major differences of our system to the most common usage of visual markers is that users take the visual marker to a fixed camera instead of actively taking a picture with a camera. Based on our results, we argue that usability is greatly improved by this design, since the manual effort spent on taking the picture and problems with resolution and focusing are removed. Issues introduced by this setup, like loss of tags, are easily solved as printing tags is straightforward and cheap. In our exploration, we additionally found that scanning is even stable for old, raddled and dirty tags.

4.3 Proximity Matters

The most critical issue deciding over use and non-use of the system seems to be the proximity of the scanning station to the desired objects. As long as the scanning station is next to the coffee machine, users operate the system in parallel to the preparation of beverages, which has a positive influence on paying behavior.

At a time when we were forced to rearrange the setup due to the renovation of the kitchen, even people passing the scanning station while returning to their office forgot about paying and later had to make up for it with the help of the web-based interface. This effect appeared even under the condition that the system was still quite close to the beverage preparation place, i.e. within 5 meters distance. We compared purchases from that period with the same period one year earlier. Although the number of average bookings per user stayed nearly constant (M=4.2 adjacent; M=4.3 detached) the average number of tag bookings decreased a lot (M=10.6 adjacent; M=6.6 detached).

Hence, proximity ensures a proper integration into the process, which is a key attribute according to our three-year deployment. When informally asking why the system was not used when it was slightly farther away, one answer was that taking and paying a drink felt like two different tasks. The subjective feeling of the time required changed and made the overall experience less pleasant. People also started to wait until they had several empty bottles on their desks before making a bulk payment using the web interface.

4.4 Simple Design Works for Groups with High Internal Trust

We found that the Barkeeper's rather simple design works because our group has a high level of internal trust and we do not have to care about complex security and privacy issues. Nevertheless, we have experienced ourselves that the system is mature enough for everyday use in the real world. Very simple features of the system like the email notification on purchases introduced enough transparency such that the system never was put into question. The reliable scanning process with an error rate close to zero certainly has also contributed to that.

4.5 Iterative and Autobiographic Design Works

The success of the system shows that it is a helpful approach to start with a very simple solution, which solves a single key problem experienced in daily life, and to evolve the system only based on concrete user suggestions. The developer is among the user group and enjoys people making suggestions for new features. Some design decisions, which may be seen questionable for an outside user (like the purely auditory feedback), evolved in a social process and therefore enjoy strong support by the group, making them a "normal" part of everyday life even for new members.

5 Conclusions

In this paper, we presented a simple but effective system for keeping track of individual beverage consumption within a group. We explained the design evolution and the long-term results of a visual marker-based system for this purpose.

We argue that visual markers are well-suited for systems which can integrate their use into the standard work process – in our case scanning the tags while coffee is brewed. Interestingly, although more time-effective alternatives were provided, users preferred to use visual markers.

Despite the basic interaction mode being slower than a traditional tally list, users were able to build up own strategies for fast usage. Combined with additional enhancements such as combi-tags, the marker-based interaction can now even outperform tally marks in terms of speed. Nevertheless, performance proved to be not the most important factor of every interactive system. Even a slower interaction may be better suited for certain real-life interaction tasks if it integrates well into the usage scenario and is appealing to the user for other reasons.

The results of our user study and the qualitative data about the system itself showed that, although the system offers various possibilities for extended user privacy and security, users do not take advantage of those. Providing a mostly error-free and easily recoverable payment system leads to a trust level that is beyond security or privacy worries. We conclude that systems providing a high transparency of their inner workings and data handling are accepted to be more privacy invasive.

Our findings can easily be applied to other domains that meet similar conditions of 1) giving individual access to global resources at different points 2) high internal trust within a user group and 3) appropriate space and interaction design that allows integrating the scanning interaction. Examples for similar situations where our design could be used are the usage of copy machines, lending out equipment or taking over routine jobs shared in a group. For usage in a public scenario such as a museum, the design is not adequate due to the missing group trust. Nevertheless, the idea of using a scanning station for movable QR-tags can also be reused in a museum (or similar public environment), e.g. by printing the QR-codes on the entrance tickets [4].

As the system's development has always been inspired by users' needs and ideas, they will definitely help to enhance it in the future. As we found out that performance is not necessarily a critical factor, features that support the playfulness of the system will most likely be added and improved. Overall, we believe the system can be used as a blueprint for a number of application scenarios in closed user groups, far beyond just keeping track of coffee consumption.

Acknowledgements. We want to thank our former colleague Raphael Wimmer for reminding us of important details; Gregor Broll for providing some excellent related work; Alexander Wiethoff for his input at the start of writing this; Anita Szász for inspiring us to create the Barkeeper; Rainer Fink for keeping the server in a good shape; Denso Wave for inventing the QR codes and finally the former members of our lab that participated in the questionnaire. Besides this we want to thank the Barkeeper for being so patient with us and to all of you: QR amazing.

References

1. Beigl, M., Gellersen, H.W., Schmidt, A.: Mediacups: experience with design and use of computer-augmented everyday artefacts. Computer Networks 35(4), 401–409 (2001)
2. Jarvis, M.J.: Does caffeine intake enhance absolute levels of cognitive performance. Psychopharmacology 110(1-2), 45–52 (1993)
3. Neustaedter, C., Sengers, P.: Autobiographical design in HCI research: designing and learning through use-it-yourself. In: Proceedings of the Designing Interactive Systems Conference, DIS 2012, pp. 514–523. ACM, New York (2012)
4. Sawers, P.: With more than 5m visitors so far, Google's London Web Lab experiments are still going strong, http://thenextweb.com/google/2013/01/13/3-million-visitors-later-googles-chrome-web-lab/ (last accessed March 22, 2013)
5. Want, R.: An introduction to RFID technology. Pervasive Computing 5(1), 25–33 (2006)
6. Want, R., Fishkin, K.P., Gujar, A., Harrison, B.L.: Bridging physical and virtual worlds with electronic tags. In: Proc. CHI 1999, pp. 370–377. ACM Press (1999)
7. Want, R.: Near field communication. Pervasive Computing 10(3), 4–7 (2011)

Moment Machine: Opportunities and Challenges of Posting Situated Snapshots onto Networked Public Displays

Nemanja Memarovic[1], Ava Fatah gen Schieck[2], Efstathia Kostopoulou[2], Moritz Behrens[2], and Martin Traunmueller[3]

[1] Faculty of Informatics, University of Lugano (USI), Lugano, Switzerland
[2] The Bartlett, University College London, London, UK
[3] UCL ICRI-Cities, University College London, London, UK
nemanja.memarovic@usi.ch,
{ava.fatah,ucftefk,moritz.behrens,
martin.traunmueller.11}@ucl.ac.uk

Abstract. Large public displays are becoming a ubiquitous resource in the urban environment. Interconnected over the Internet these hitherto isolated "ad displays" could become a novel and powerful communication medium – networked public displays. One example for such a novel type of communication is their use as community tools. Scattered across the urban landscape and equipped with additional sensors, such as cameras, they provide the opportunity for local community members to take images of themselves and leave their "mark" in the setting, e.g., on their way to school, work, or meeting with friends. In order to understand the potential of posting situated snapshots on networked public displays in the context of place-based communities we designed and developed the Moment Machine – a networked public display application that allows one-click photo capture. In this paper we report on identified opportunities and challenges emerging from 6 user trials in the wild at 2 locations.

Keywords: networked public displays, urban screens, situated snapshots, community interaction, public space, urban computing, urban informatics.

1 Introduction

The urban landscape is getting "painted" with large public screens: from streets across small cafes to building facades we can find digital displays showing street maps, restaurant menus, or latest brand advertisement [10]. Although most of public displays represent singular installations that run locally stored power points or images, it is not hard to imagine that they will soon be connected and networked over the Internet [3, 15, 16]. Networked and empowered with additional sensors such as cameras networked public displays will constitute a novel and powerful communication medium for the 21st century [3].

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 595–602, 2013.

Dispersed across the urban landscape and embedded in the setting, they offer an opportunity to connect local community members, e.g., by having them support picture taking and sharing across the display network. For example, a school kid on his/hers way to school could be capturing images with his friends on a network of public screens, a group of teenagers could take photos while they are out-and-about on a Friday night, while adults with jobs could be taking photos to and from their way to work.

By taking images and leaving them in the settings (1) community members will express themselves and will leave their "mark". In return the photos taken by people in display locality could (2) stimulate community awareness by providing information on "who is around". They could also (3) enrich their locality and provide insights into other locations through photos captured in different environments, thus stimulating one of the human needs in public spaces: the need for discovery of new features and places [12].

As part of an ongoing "Screens in the Wild" research project [18] and in order to better understand the implications of posting situated snapshots onto networked public displays in the context of place-based communities we designed and developed Moment Machine – a networked public display application that allows one-click on-display photo capture. The Moment Machine is described in detail in section 3, right after the related work section (section 2). Our study design is summarized in section 4 where we describe six user trials "in the wild" at 2 locations where we conducted 55 interviews with passers-by. We present findings from our study in section 5, organized around (1) possible motivations for posting situated snapshots onto networked public displays, (2) discussion on where the images could be displayed once they are taken, and (3) challenges in promoting and sustaining their use. Finally we provide concluding remarks in section 6.

2 Related Work

Research on networked public displays has seen its pioneering work in the early 80s where one of the prominent examples from that period is the 'Hole in Space' [5], a project that connected two urban spaces in New York and Los Angeles through a video link. Connected Urban Spaces [4] and Telectroscope [20] are examples of similar and more recent projects.

In more recent years work in the area has shifted from creating simple video links between two places into creating more interactive multimedia applications [2, 13] and how they can enrich urban spaces and promote community interaction and place awareness [11]. For example, Alt et al. [1] investigated how networked public displays can support exchange within urban communities by creating a digital version of traditional/analog notice boards.

The work of Ojala et al [16], Hosio et al. [6], North et al. [15], Peltonen et al. [17], and Taylor and Cheverst [19] falls closest to the work described in this paper as they have investigated how images can be used on networked public displays. Ojala et al.'s UBI-Postcard allows passers-by to take photos using a camera attached to a display.

Once the photos are taken users can send the photo to an email address. Hosio et al.'s Ubinion similarly used public displays as input devices: young adults/teenagers could use them to create posters with speech bubbles that would represent their concerns about the city of Oulu. Posters created on displays (i.e., a network of displays) would be posted on dedicated Facebook and Twitter accounts. While UBI-Postcards and Ubinion used displays as input devices, Taylor and Cheverst's Wray Photo Display showed images uploaded from a dedicated website (images were uploaded by Wray villagers). Similarly, Peltonen et al.'s CityWall display showed user-contributed images from Flickr tagged with a specific tag. In the same manner North et al.'s ScreenGram shows images from Instagram with a predefined hashtag on a display network.

While all the above work used public displays for either taking the images [6, 16] or showing them on networked displays [15, 19] the study described here is the first one that used networked public displays for both. Also, our work complements the current body of research on networked public displays by (1) describing possible motivation for taking and posting situated snapshots onto public display networks, (2) discussing where could the images be displayed once they are taken, as well as (3) summarizing challenges in promoting and sustaining their use.

3 The Moment Machine

The Moment Machine is a networked public display application that allows passers-by to simply capture their everyday moments by taking an image through a display and spreading it across the network. The Moment Machine's user interface is shown in Fig. 1-a. Its simple user interface was inspired by previous research that showed that passers-by do not spend a lot of time looking at public displays [7] and that live video feed represent a good mean for getting their attention [14]. For this reason we show a live video feed (1 in Fig. 1-a) and allow passers-by to take an image by simply pressing a button (3 in Fig. 1-a). Users can also change the "look" of their snapshot by selecting a filter before they take a photo (2 in Fig. 1-a), somewhat similar to the popular social media application Instagram [9]. In order to give the passers-by enough time to prepare themselves, image capture is delayed by five seconds, which is indicated through a countdown timer appearing instead of the "hands" button (2 in Fig. 1-a). Once the moment is captured users have thirty seconds to decide if they want to leave the moment on a display (shown in Fig. 1-b). After that the moment appears on the screen and across the display network (4 in Fig. 1-a). Passers-by also have the possibility to browse through moments captured at all locations (4 in Fig. 1-a).

4 Study Design

In order to uncover the potential of posting situated snapshots onto networked public displays we conducted six user trials "in the wild" at 2 locations in London, i.e., Walthamstow and Leytonstone. Both locations had a display in a public urban setting provided by the "Screens in the Wild" project [18] and were connected thus making a small display network ideal for technology probing [8]. User trials were conducted between 12AM and 5PM on working days. They lasted between 2.5 and 4.5 hours and partially overlapped. In total we conducted of 7.5 hours of trials at Leytonstone and 10.5 hours of trials at Walthamstow. Because the goal of the study was to “probe” and understand the implications of taking situated snapshots on local community members we decided to invite users to interact with the Moment Machine.

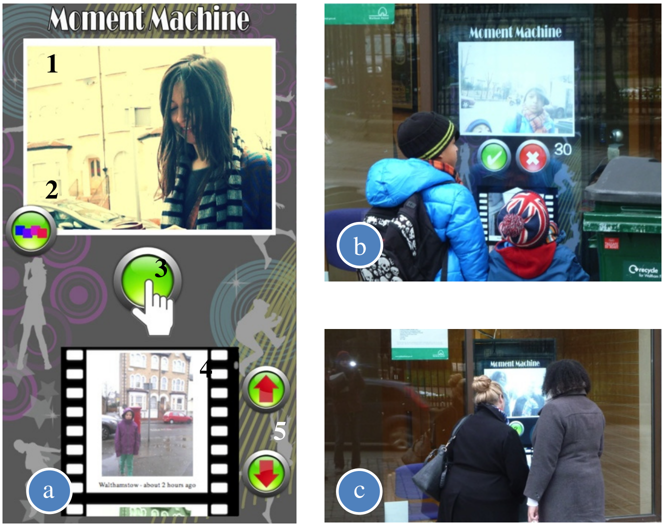


Fig. 1. The Moment Machine: a) user interface and b-c) passers-by interacting with it

During the trials we conducted 55 interviews (27 individual and 18 group interviews) with 71 people who interacted with the Moment Machine. At each location there were between 1 and 3 researchers who invited passers-by to interact with it. Passers-by were asked to take a photo and browse through existing ones on the display. After taking the snapshot researchers conducted semi-structured interviews and asked the participants about their general impression of the Moment Machine, e.g., would they come back to take more photos and browse through the ones shown on a display, and where they would like to see the images appear. After conducting the interview researchers wrote down the notes with some verbatim.

5 Findings

After collecting all the data we conducted an affinity diagram analysis and categorized the data, focusing on (1) possible motivations for taking images through networked public displays and leaving them in urban settings. We were also interested in (2) uncovering other digital places where the images could appear, e.g., Facebook or personal email as well as (3) understanding the challenges in promoting and sustaining the use of the Moment Machine application.

Each interview received an identifier comprising from the first letter of the location (L – Leytonstone, W – Walthamstow) and a sequential number of the interview, e.g., L01 would be the first interviewed person/group from Leytonstone while W04 would be the fourth interviewed person/group from Walthamstow. Considering the qualitative nature of our study our results are not intended as statistical evidence for expected behavior. Instead our results serve as an input to the design process by identifying potential tendencies of passers-by.

We first describe possible motivation for taking situated images and effects of posting them in urban settings, then we discuss where could the images taken through networked public displays appear, and in the end we present some of the challenges in sustaining their use.

5.1 Motivations and Effects

One of the possible reasons why passers-by would take images and leave them on public displays would be to "capture everyday life" (L04). For example, L06 stated that he would simply take pictures while going to or coming back from his work. Similar comment was received from L12 who said that he would take images on his way to and from school. Although we ran six user trials "in the wild" on arbitrary days this type of behavior was somewhat confirmed as we did get 3 returning users (L12, W18, W19) who brought additional participants to the display.

Taking images of oneself alone or in company of friends and family is another reason (W12, W13). This is also reflected with types of images people would like to see on public displays, i.e., images of people they know and have seen in their locality. As W12 points out "It is interesting if there is a photo of someone I know." As mentioned previously, the Moment Machine was tested in settings where displays were already running scheduled content. As interviewees from L01 point out "[we] will now pay more attention to the screen… [because we] expect to see images of people from Leytonstone [where they are residents] and our friends from Walthamstow." Similar comments were received from W31 who was commenting on people in the images "I know him… he lives there… he has a store…" Seeing images of people from the locality can bring up community awareness, i.e., who is around the neighborhood. This was explicitly stated by L05 and L13. However this type of communication can go even further and target individuals. For example, when capturing the moment W29 made a heart symbol with his hands. When asked why he did it he said that it is a sign between him and his girlfriend who was at work.

Another reason for taking the images would be getting "5 minutes of fame". This is best captured by tongue-in-cheek shouting from two young girls "We are going to be famous!" Similarly, while a girl from L19 was taking a picture a man passing-by stated "Look, you're famous!" when he saw her picture on a display. Also, L16 stated that she likes that people can see her, while W17 was fascinated that people can see him at Leytonstone. Another benefit from capturing the moments outside would be to actually spend time in the physical environment. As L13 points out "Kids these days spend too much time in front of the computer, this might get them to play outside."

Mashing of images/moments from the 2 locations can lead to universal connections, or “universality” as stated by L15. She liked the universality of the machine and that everyone can get connected and mashed through the Moment Machine.

In general people were interested in seeing images from the “other” location, thus reflecting the need for discovery of new features and places [12]. For example, the two people from L01 showed their interest in the Moment Machine application because they were from Leytonstone and had friends in Walthamstow. However, W11 from Walthamstow said that “[I] know the place [Leytonstone]. I don’t know the people who live there.” Because he did not know people from Leytonstone he was indifferent to seeing images from it. This shows that displays could connect to locations where users have friends/family/acquaintances, e.g., by harvesting a Facebook profile form users in the display vicinity.

While some people wanted to get more information from the other location, e.g., a live video feed (L09), others (W29) were concerned with this stating “I don’t want just to spy on others through a website which connects me to cameras” and favored the asynchronous connection through images which allowed them to browse on their own.

5.2 Where Should the Images Appear

While most of the participants were comfortable to take the photo and leave it on displays, some wanted to know where will the image appear before taking the photo asking “Is it gonna stay here?” (L17, W14)

When we asked the participants where else would they like to see the images we received two types of answers: (1) either seeing them on a Facebook page, email, or dedicated website (18/26 participants) or (2) leaving the images to stay on the display (8/26 participants). Some of the participants did not see the point of leaving the image on the network unless they could “take” the moment with them. This is best captured by L14 who asked for her image to be removed since she could not take it with her. Interesting insight is also offered by L17 who wanted to get a printout of her photo – this was preferred method over email. In essence this finding points out that networked public screens maybe need to be better connected with users “communicative ecology” [13], i.e., with current information and communication technologies such as Facebook and email.

On the other hand we also had participants who were satisfied with having the images appearing across the network, i.e., they did not want them to appear anywhere else: “absolutely not” stated L15 when asked if the images should appear somewhere else. She liked that the images are staying at the “machine” as they are “just capturing moments of everyday life.”

5.3 Challenges in Promoting and Sustaining Use

One of the most prominent challenges in having people interested in interacting with the Moment Machine was scalability, i.e., the number of displays deployed in the urban space that run it (L07, W13, W34, W36). For example W13 points out that there are only two locations now that have this type of application while W36 stated that networked screens have to have a bigger scope and purpose.

Some of the challenges are posed by the urban environment that changes during the day. The research team noticed that when it was too bright it was hard to notice the video feed and participants had difficulties in understanding what is wrong and where they should position themselves to take the photo. Additional problems with positioning can be caused by camera placement. For example, some of the participants at Walthamstow had problems in positioning themselves to take a desired snapshot, e.g., a portrait. The camera used for photo taking was attached to the right side of the screen so if participants wanted to take a portrait they would have to look at the camera. However, some participants found this difficult and looked at the live video feed instead. Similar problems can be caused by the positioning of the screen itself. In order to make the user interface reachable to kids we intentionally made the buttons a bit lower. At Leytonstone the display is positioned a bit higher and both kids and adults did not report any problems interacting with it. However, W31 at Walthamstow commented that it is not comfortable to bend down in order to browse the photos.

Although the purpose of the application was clear to most of the participants, some had difficulties and questions regarding its purpose (L09, W26, W31, W34). A possible solution for this challenge would be to have a more obvious message on ta display, e.g., “Capture your daily moment” or “Strike a pose for the world!”

Finally some of the participants mentioned that the Moment Machine could benefit from adding new features (L09, W31, W33, W34). Some of the suggested features are relatively easy to add, e.g., adding the number of overall images that were taken (W31) while others are more sophisticated, e.g., adding image manipulation and ability to play with the background image (L09, W34). Adding new features could lead to users staying interested in the application, as suggested by Memarovic et al. [12].

6 Conclusions

In this study we used technology probe based approach to understand how posting situated snapshots on networked public displays could be beneficial for place-based communities. For this purpose we designed and developed the Moment Machine application and conducted 6 user trials “in the wild” at 2 locations in London. Our findings from observing 71 people interacting with the Moment Machine and conducting 55 interviews are promising and show that networked public displays can be used by local community members to capture images “on the go.” In this paper we reported on (1) potential motivation for and effects of posting situated snapshots onto networked public displays, (2) where the images could be displayed once they are taken, and (3) challenges in promoting and sustaining their use.

Acknowledgments. The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 244011, the EPSRC-funded project “Exploring the potential of networked urban screens for communities and culture” under the grant agreement no. EP/I031839/1, with the support of UCL VEIV center. We would also like to thank professor Marc Langheinrich for his useful comments on the early version of the paper and “Screens in the Wild” members for invaluable contribution.

References

1. Alt, F., et al.: Digifieds: insights into deploying digital public notice areas in the wild. In: Proceedings of MUM 2011, pp. 165–174. ACM, New York (2011)
2. Clinch, S., Davies, N., Friday, A., Efstratiou, C.: Reflections on the long-term use of an experimental digital signage system. In: Proceedings of the 13th International Conference on Ubiquitous Computing, pp. 133–142. ACM, New York (2011)
3. Davies, N., Langheinrich, M., José, R., Schmidt, A.: Open Display Networks: A Communications Medium for the 21st Century. IEEE Compt. 45(5), 58–64 (2012)
4. Fatah gen. Schieck, A., Fan, S.: Connected urban spaces. In: Proceedings of SSS8 (2012)
5. Hole in Space, `http://bit.ly/de6vOG`
6. Hosio, S., Kostakos, V., Kukka, H., Jurmu, M., Riekki, J., Ojala, T.: From School Food to Skate Parks in a Few Clicks: Using Public Displays to Bootstrap Civic Engagement of the Young. In: Kay, J., Lukowicz, P., Tokuda, H., Olivier, P., Krüger, A. (eds.) Pervasive 2012. LNCS, vol. 7319, pp. 425–442. Springer, Heidelberg (2012)
7. Huang, E.M., Koster, A., Borchers, J.: Overcoming assumptions and uncovering practices: When does the public really look at public displays? In: Indulska, J., Patterson, D.J., Rodden, T., Ott, M. (eds.) Pervasive 2008. LNCS, vol. 5013, pp. 228–243. Springer, Heidelberg (2008)
8. Hutchinson, H., et al.: Technology probes: inspiring design for and with families. In: Proceedings of CHI 2003, pp. 17–24. ACM, New York (2003)
9. Instagram, `https://en.wikipedia.org/wiki/Instagram`
10. Kostakos, V., Ojala, T.: Public Displays Invade Urban Spaces. IEEE Pervasive Computing 12(1), 8–13 (2013)
11. Memarovic, N., Langheinrich, M., Alt, F.: The interacting places framework. In: Proceedings of PerDis 2012. ACM, New York (2012)
12. Memarovic, N., Langheinrich, M., Alt, F., Elhart, I., Hosio, S., Rubegni, E.: Using Public Displays to Stimulate Passive Engagement, Active Engagement, and Discovery in Public Spaces. In: Proceedings of MAB 2012, pp. 55–64. ACM, New York (2012)
13. Memarovic, N., Langheinrich, M., Rubegni, E., David, A., Elhart, I.: Designing "interacting places" for a student community using a communicative ecology approach. In: Proceedings of MUM 2012. ACM, New York (2012)
14. Müller, J., Walter, R., Bailly, G., Nischt, M., Alt, F.: Looking glass. In: Proceedings of CHI 2012, pp. 297–306. ACM, New York (2012)
15. North, S., Schnädelbach, H., gen Schieck, A.F., Motta, W., Ye, L., Behrens, M., Kostopoulou, E.: Tension Space Analysis: Exploring Community Requirements for Networked Urban Screens. In: Kotzé, P., Marsden, G., Lindgaard, G., Wesson, J., Winckler, M. (eds.) INTERACT 2013, Part II. LNCS, vol. 8118, pp. 81–98. Springer, Heidelberg (2013)
16. Ojala, T., Kukka, H., Lindén, T., Heikkinen, T., Jurmu, M., Hosio, S., Kruger, F.: UBI-hotspot 1.0. In: Proceedings of ICIW 2010, pp. 285–294. IEEE CPS (2010)
17. Peltonen, P., Salovaara, A., Jacucci, G., Ilmonen, T., Ardito, C., Saarikko, P., Batra, V.: Extending Large-Scale Event Participation with User-Created Mobile Media on a Public Display. In: Proceedings of MUM 2007, pp. 131–138. ACM, New York (2007)
18. Screens in the Wild, `http://screensinthewild.org/`
19. Taylor, N., Cheverst, K.: Supporting Community Awareness with Interactive Displays. IEEE Compt. (Special Issue on Interactive Digital Signage) 45(5), 26–32 (2012)
20. Telectroscope, `http://www.telectroscope.net/`

Overview Scrollbar: A Scrollbar Showing an Entire Document as an Overview

Ko Mizoguchi, Daisuke Sakamoto, and Takeo Igarashi

The University of Tokyo
{poohtokou,sakamoto,takeo}@is.s.u-tokyo.ac.jp

Abstract. A scrollbar is the most basic function of a graphical user interface. It is usually displayed on one side of an application window when a displayed document is larger than the window. However, the scrollbar is mostly presented as a simple bar without much information, and there is still plenty of room for improvement. In this paper, we propose an overview scrollbar that displays an overview of the entire document on it and implemented four types of overview scrollbars that use different compression methods to render the overviews. We conducted a user study to investigate how people use these scrollbars and measured the performance of them. Our results suggest that overview scrollbars are more usable than is a traditional scrollbar when people search targets that are recognizable in overview.

Keywords: user interface, scrollbar, document navigation.

1 Introduction

People can access many documents easily by using computers, but they also face various problems inherent to electronic devices, so there is strong demand for improvements to electronic document interfaces. As an example of one such improvement, we redesigned the scrollbars of the computer application window. Traditional scrollbars are drawn compactly on one side of applications in order to not reduce the main document space. However, since the resolution of displays is getting higher nowadays, users would be able to see main documents in a comfortable size without the need for making scrollbars smaller.

In this paper, we propose an overview scrollbar, which is a wide scrollbar that shows an overview of an entire document. Users can browse through a document by using an overview of the document, which is placed on one side of the application window, and the scrollbar enables the users to quickly grasp an overview of a full document and jump to any location immediately by clicking the overview. We implemented four types of overview scrollbars (uniform scaling, vertical compression, fisheye, and perspective) that use different methods to compress the overviews. The first two scrollbars compress the overviews by using a constant ratio. When a document is very large, the compression ratio gets very high, so recognizing detailed contents from the overview may become difficult. The latter two scrollbars compress the overviews by spatially varying the ratios with focus and context areas.

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 603–610, 2013.

The compression ratio is low in the focus area and high in the context area. We conducted a user study to investigate the usability and the problems of the overview scrollbars. Results show that they were more usable than were the traditional scrollbars when the search targets can be recognized only by overviews; however, we found some problems, especially with the fisheye and perspective ones.

2 Related Work

Shneiderman suggested basic improvements to scrollbars in 1992, such as placing marks on scrollbars or drawing page numbers on sliders [1]. On the basis of these improvements, much research was done afterwards [2,3,4,5], but most of it was simple functions added to traditional scrollbars and did not mention drastic modifications such as our wider scrollbar.

Document visualization research has been done in the field of HCI. The *Reader's Helper* [6] is a document navigation system that uses *Thumbar*, which shows an overview of an entire document. Likewise, the *Popout Prism* [7] uses an overview panel to do so. These pieces of research used interfaces similar to our overview scrollbars, but they did not examine how to scale the overview in depth. Thus, we examined the scaling methods that are used in these pieces of research with our novel methods.

Hornbæk et al. conducted an experiment to compare the usability of linear, fisheye, and overview+detail interfaces on a reading environment [8,9]. Baudisch et al. implemented *Fishnet*, which is a web browser that uses a fisheye view [10], and compared the usability between that and other interfaces. The purpose of the former study was to investigate which interfaces were suitable for enhancing a reader's degree of comprehension, and the purpose of the latter was to investigate which interfaces perform well when readers search text based targets in a document by using highlighted annotations. In this paper, we investigate the performance of scrollbars when readers search targets in a large document by using its full overview.

3 Overview Scrollbar

We propose four types of scrollbars showing document overviews — uniform scaling, vertical compression, fisheye, and perspective. The area users see in the main view is highlighted on the scrollbar. Users can scroll the document to any location by left clicking or dragging on the scrollbar. Although the methods for compressing the overviews are different in these scrollbars, there are basic rules in compression.

The width of the overview scrollbar is limited to 20% of the application window. When the height of the document is too short, the overview is scaled so that the width fits its maximum size, preserving the aspect ratio. In this case, the height of the overview is shorter than that of the application, so the overviews are arranged on the top and nothing is drawn under the overviews. In the case that the height of the document is too large, the overview is compressed as follows.

(a) **Uniform Scaling Scrollbar:** Figure 1(Left) and Figure 1(a) show the uniform scaling scrollbar. The aspect ratio of the overview is maintained regardless of the vertical size of the document. Thus, if the vertical size of the document is too large, the width of the overview may become too narrow.
(b) **Vertical Compression Scrollbar:** Figure 1(b) shows the vertical compression scrollbar. When the vertical size of the document is too large to maintain the aspect ratio of the overview, the overview is simply compressed vertically so that the height matches that of the application.
(c) **Fisheye Scrollbar:** Figure 1(c) shows the fisheye scrollbar. When the overview needs to be compressed vertically, a focus lens appears whose height is 70% that of the application. Within the lens, the compression ratio is low and constant with the preserved aspect ratio but gets higher as the distance from the lens increases. In this scrollbar, the width outside of the lens is compressed as well as the height. The user has to control two entities in this scroll bar: the focus lens and the knob (position of the contents shown in the main view). Users respectively move the knob and the lens by holding down the left or right buttons and dragging with the mouse. The user can also move them simultaneously by dragging with both mouse buttons held down.
(d) **Perspective Scrollbar:** Figure 1(d) shows the perspective scrollbar. This interface is equivalent to the fisheye scrollbar except that the width outside of the lens has a constant size, which is 20% of the application.

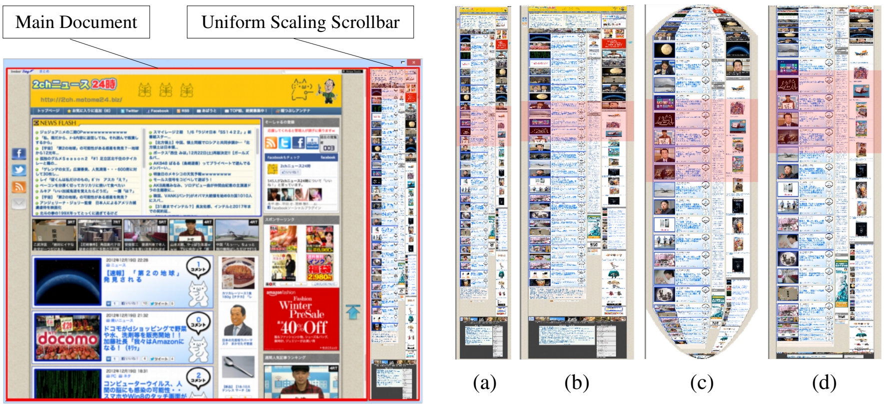


Fig. 1. Vertically large document with the uniform scaling scrollbar and the overview scrollbars: (a) uniform scaling, (b) vertical compression, (c) fisheye, and (d) perspective

4 User Study

We conducted a user study to investigate the performance and problems of the traditional scrollbar (called "standard scrollbar" below) and the overview scrollbars. We measured the task completion time, the distance scrolled by dragging the scrollbar, and the distance scrolled with the mouse wheel. Nine students participated in this study. All participants were computer science department students who were familiar with the standard scrollbar and had never used the overview scrollbar.

Fig. 2. Sample task execution screens. Search targets are surrounded by green rectangles. (a) Image search task, (b) Product search task, (c) Round trip task, and (d) Source code search task.

4.1 Tasks

Participants performed four types of tasks on each scrollbar. The same types of tasks were created so that they would all have the same of difficulty. All tasks required the participants to search for and click objects on a vertically long document. The documents were created beforehand as an image by processing the screen captures of real websites or source code.

(a) **Image search task:** Participants were asked to search an image on a large document (Figure 2(a)). The document size was about 1000 × 7000 pixels and had about 60 images on it. Before the task began, we specified an image of about 190 × 140 pixels, whose instances appeared only once in the document.

(b) **Product search task:** Participants were asked to search for a product on a large document (Figure 2(b)). The document size was about 1000 × 6500 pixels and had 60 products on it. The products were arrayed in 3 columns and 20 rows and had one image representing the product and its title below the image. Before the task began, we specified a product that is unique in the document by using text. The text did not contain product specifications such as maker, model, or name but visual features such as color or form.

(c) **Round trip task:** Participants were asked to search for two links on a large document (Figure 2(c)). The document size was about 1000 × 6000 pixels and was mainly composed of text. There were two links highlighted in red and located near the top and bottom of the document. The participants were required to click these links six times mutually scrolling the documents up and down repeatedly.

(d) **Source code search task:** Participants were asked to search for a specified line on a large source code (Figure 2(d)). The source code contained approximately 300 lines written in Java. We chose words that appeared roughly 20 times in the code and highlighted them in yellow. The participants were asked to locate a line in which the word was used in a specific way, for example, a line where the highlighted word is substituted by a value or a word that appears in a conditional "if" statement.

4.2 Questionnaire

Participants were asked to answer a questionnaire after completing all trials. The questionnaire consisted of four Likert Scale questions (5 pts.) and three freeform questions (Table 1). In the Likert scale questionnaire, the participants were asked to answer a questionnaire item by using a five point scale (5 was the most positive).

Table 1. Questionnaire

Type	#	Question
Likert	1	I felt that it was easy to get used to this method (get-used-to).
	2	I thought that this method was easy to use (easy-to-use).
	3	I thought that this method was suitable for displaying documents (suitability).
	4	I want to use this method in practice (want-to-use).
Freeform	5	Please write any good points and bad points in using this method.
	6	Please write any improvements for this method.
	7	Please write anything on your impressions or findings.

4.3 Results

Performances

A one-way within-subject ANOVA was conducted to compare the scrollbars. In the distance scrolled by dragging the scrollbar and with a mouse wheel, there were no statistical differences between all of them. However, there were significant differences in the task completion time. Figure 3(a) shows the average completion times for each task. In the round trip task, there were significant differences between the scrollbars ($F = 12.75$, $p < .01$). The vertical compression, uniform scaling, and perspective scrollbars were significantly faster than was the standard one ($p < .05$). Likewise, the uniform scaling and the vertical compression ones were significantly faster than was the fisheye one ($p < .05$), and the vertical compression one was significantly faster than was the perspective one ($p < .05$). In the image search task, there were no significant differences. However, there was a tendency for the standard scrollbar to be faster than any of the others. In the source code search task, two participants got lost in the source code for a long time when using the standard and perspective scrollbars. Therefore, the average completion times for these two seemed very slow, but there were no statistical differences.

Preferences

A one-way within-subject ANOVA was conducted in the questionnaire. Figure 3(b) shows the average scores of the participants for each task. In the first question, there were significant differences between the scrollbars ($F = 4.33$, $p < .01$). In pairwise comparisons, the uniform scaling scrollbar had a significantly higher score than did the fisheye one ($p < .05$). In the second question, we also found significant differences between the scrollbars ($F = 3.54$, $p < .05$). In pairwise comparisons, the vertical compression scrollbar had a significantly higher score than did the standard one ($p < .05$). In the third and forth questions, there were no significant differences. However, the standard scrollbar tended to be rated lower than any of the others.

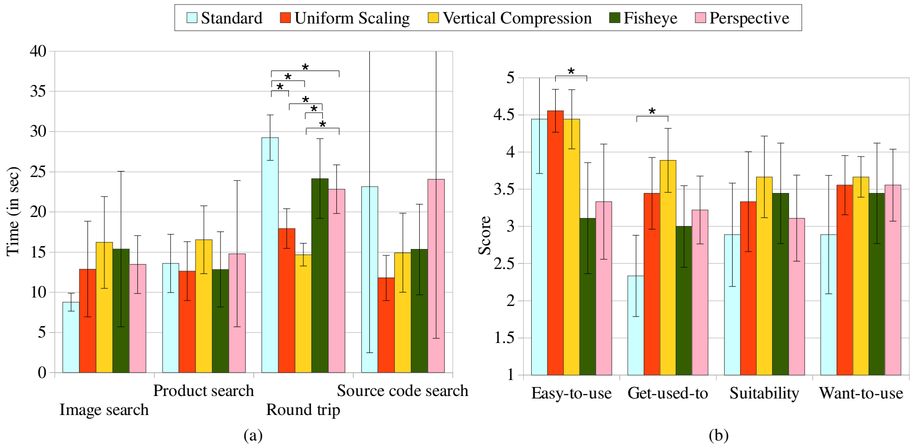


Fig. 3. (a) Average task completion times, (b) Average scores (error bars indicate 90% confidence intervals.)

Freeform Comments

We collected several comments for each scrollbar. For the standard scrollbar, some participants said, *"It was very unstable to search for something in a document."* We considered that they could not grasp an overview of the documents immediately, and they could not see where the targets were likely to be placed.

Some participants said, *"I felt little difference between the uniform scaling and vertical compression scrollbars."* The round trip task was easy with both of the scrollbars because participants could see where the red links were placed in one glance by looking at the overview even if it was highly compressed. However, most participants noticed that identifying detailed contents only by seeing the overview was difficult. Thus, in the image search and the product search tasks, these scrollbars might have been used simply as a wide scrollbar.

For the fisheye and perspective scrollbars, some commented that the images within the focus area were relatively easy to recognize, although it seemed to still be difficult to recognize contents completely. Likewise, almost all of the participants had difficulties in getting used to these scrollbars. They said, *"The method of moving the focus lens by right-clicking was not intuitive and difficult to get accustomed to,"* and some of them commented that it would be better to integrate the right button function into the left button.

Observations

We observed how the participants performed the tasks. In the image search and product search tasks, most participants looked at the overview first and tried to find the target objects only from the overview. However, once they noticed that they could not recognize the contents only from the overview, they started to search targets mainly by looking at the main document. In the end, they used the overview scrollbars only for scrolling (although there were a few participants who could search images only from the overview when using the fisheye and perspective scrollbars.) Furthermore, we found that many participants seemed to have difficulties in getting used to the fisheye and perspective scrollbars. The successful participants tended to use these scrollbars by clicking the left and the right mouse buttons simultaneously, moving the focus lens and scrolling the document at the same time. They seemed to search for the targets efficiently.

4.4 Summary of Results

Our results suggest that there were significant differences in the average completion times in the round trip task. In this task, we found that the overview scrollbars were significantly faster than was the standard one. When the participants performed this task, unlike any other tasks, they were not required to recognize the detailed contents. Because of this feature, the advantages of the overview scrollbars that the participants can grasp an overview of the documents immediately and jump to the location they wanted immediately might have been emphasized. However, there were little differences except for in the round trip task. We guessed the reason for this result by checking the feedback and by observing the participants. Many participants suggested that the overview scrollbars (even the fisheye and perspective ones) were not powerful enough to let them completely grasp the detailed contents in a document. Therefore, as we observed, the participants tended to cope with the tasks by simply seeing the main document after a short trial of identifying the contents only with the overviews. Or rather, such a meaningless trial might produce some delays in image search and product search tasks, which might be one of the reasons why the standard scrollbar seemed to perform better.

5 Conclusion

We presented four types of overview scrollbars, conducted a user study, and analyzed study results. The result of the study and the feedback from the participants revealed to us many aspects for improvement in usability and for solving the problems of the overview scrollbars. First, we think that uniform scaling and vertical compression are already acceptable methods for document navigation systems because they marked a relative superior result to the other scrollbars both in performances and evaluations. Second, the user interfaces of the fisheye and perspective scrollbars were not perfect. If these two are implemented in a more sophisticated way, these methods might become more useful interfaces for navigating documents.

However, there are some limitations on this research. First, our user study was conducted in a small sample group so that a formal user study is necessary to verify usability of the interface. Likewise, comparison studies between not only the scrollbars but also the scrollbars and the other types of methods that were useful when showing long documents such as *Perspective-Drag* or *Zoom-and-Drag* mentioned in [11]. Finally, we didn't use eye-tracking systems to analyze the user's behavior through the study. How users switch visual attentions between scrollbars and contents and how they interact with the scrollbars are both interesting. By quantifying these interactions, there may arise valuable results and discussions.

References

1. Shneiderman, B.: Designing the user interface: strategies for effective human-computer interaction. Addison-Wesley Longman Publishing Co., Inc., Boston (1986)
2. Björk, S.: The scrollsearcher technique: Using scrollbars to explore... In: Proc. of Interact 2001, Eighth IFIP TC.13 Conference on Human-Computer Interaction, pp. 745–746. IOS Press (2001)
3. Byrd, D.: A scrollbar-based visualization for document navigation. In: Proceedings of the Fourth ACM Conference on Digital Libraries, DL 1999, pp. 122–129. ACM, New York (1999)
4. Alexander, J., Cockburn, A., Fitchett, S., Gutwin, C., Greenberg, S.: Revisiting read wear: analysis, design, and evaluation of a footprints scrollbar. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2009, pp. 1665–1674. ACM, New York (2009)
5. Laakso, S.A., Pekka Laakso, K., Saura, A.J.: Improved scroll bars. In: Proceedings of Human Factors in Computing Systems, CHI 2000, pp. 97–98. ACM Press (2000)
6. Graham, J.: The reader's helper: a personalized document reading environment. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 1999, pp. 481–488. ACM, New York (1999)
7. Suh, B., Woodruff, A., Rosenholtz, R., Glass, A.: Popout prism: adding perceptual principles to overview+detail document interfaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2002, pp. 251–258. ACM, New York (2002)
8. Hornbæk, K., Frökjær, E.: Reading of electronic documents: the usability of linear, fisheye, and overview+detail interfaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2001, pp. 293–300. ACM, New York (2001)
9. Hornbæk, K., Frökjær, E.: Reading patterns and usability in visualizations of electronic documents. ACM Trans. Comput.- Hum. Interact. 10(2), 119–149 (2003)
10. Baudisch, P., Lee, B., Hanna, L.: Fishnet, a fisheye web browser with search term popouts: a comparative evaluation with overview and linear view. In: Proceedings of the Working Conference on Advanced Visual Interfaces, AVI 2004, pp. 133–140. ACM, New York (2004)
11. Guiard, Y., Du, Y., Chapuis, O.: Quantifying degree of goal directedness in document navigation: application to the evaluation of the perspective-drag technique. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2007, pp. 327–336. ACM, New York (2007)

Participatory Design for Cultural Representation: A Cultural Transparency Perspective

Amalia G. Sabiescu[1] and Nemanja Memarovic[2]

[1] New Media in Education Laboratory, Faculty of Communication Sciences, Università della Svizzera italiana, Via G. Buffi 13, 6900 Lugano, Switzerland
[2] Faculty of Informatics, Università della Svizzera italiana, Via G. Buffi 13, 6900 Lugano, Switzerland
{sabiesca,nemanja.memarovic}@usi.ch

Abstract. Participatory design approaches are being increasingly employed for designing digital artefacts and information systems with and for local communities. These cases require a reconceptualization of PD processes to account for widened knowledge gaps between designers and community members, and new patterns of community-defined design goals. In this paper we provide a perspective on the design process that will help designers to better plan their involvement in participatory projects with local communities. Our analytical stance resides on an interpretation of Étienne Wenger's theory of cultural transparency. Participatory design is analysed as an iterative process of decoding and encoding that involves users/local people and designers having as outcome understanding (through decoding) and representations (through encoding). *Cultural transparency*, achieved when the two agents advanced sufficient understanding on the other's practices, is the landmark for effective design. The paper argues for the importance of working towards attainment of cultural transparency in community-based projects, in particular when the goal is to create culturally representative artefacts. Examples of activities and suggestions for advancing cultural transparency in these contexts are provided.

Keywords: participatory design, cultural transparency, cross-cultural design, local communities, cultural representation.

1 Introduction

Participatory design (PD) approaches, traditionally used in work environments [1], are being increasingly employed for designing digital artefacts with and for local communities, serving goals defined by community interests, needs, and aspirations. For example, ICTs can be instrumental for preservation and transmission of local knowledge among different generations of the same community [2, 3], recovery of memory by appropriation and digital representation of community artefacts [4], and networking among disenfranchised indigenous communities [5]. These goals can be supported by the design and development of various information spaces, communication artefacts, and repositories, ranging from digital archives to community websites. When using PD in community contexts, three inter-related sets of challenges arise:

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 611–618, 2013.

1. The *knowledge gap* between designers and communities is widened. For designers, community settings are characterized by peculiar social and cultural features that need to be understood for creating relevant artefacts [3]. For community members, digital technology and design workflows, processes and terminology are replete with unknowns that make their informed participation difficult.
2. A reconceptualization and renegotiation of established *design methods and techniques* is required for accommodating the participation of community members that may lack digital literacy or even be illiterate. As Winschiers [6] argues, even widely accepted PD methods, such as Future Workshop, may be unfit in certain community contexts.
3. The formulation of design goals needs to take into account *relations between the features of the digital artefact* and the community's *long-term goals and bounding cultural protocols*. For example, Christen [4] reports on the design of an indigenous archive in which different levels of access were defined reflecting restricted access to local knowledge for different community groups.

This paper outlines a theoretical perspective that can shed light on the factors that are brought to bear when employing PD in community contexts. Drawing on an interpretation of Étienne Wenger's theory of cultural transparency [7,8], the paper will provide a conceptual mapping of the design continuum highlighting the dynamics of interaction between users and designers in the creation of novel artefacts. PD is analysed as an iterative process of decoding and encoding involving users and designers having as outcome understanding (through decoding) and representations (through encoding). *Cultural transparency*, achieved when the two agents advanced sufficient understanding on the other's practices, is the landmark for effective design. The paper argues for the importance of working towards attainment of cultural transparency in community contexts, in particular when the goal is to create representative artefacts.

2 Background and Related Work

The outline of the theoretical perspective exposed herein was shaped by considerations around a participatory project that aimed to explore the potential of ICTs for giving voice to minority groups starting from the grassroots [9,10]. The project, titled *Romani Voices* involved two Romani communities in rural Romania. The methodological approach employed placed an emphasis on emergence and people participation in design at two levels: (1) Activity design: people's views were integrated in the design of a content production experience for documenting local cultural practices and collective priorities; and (2) Product design: people participated in designing the information architecture and deciding upon suitable content to be included in a community website for each group involved (www.romanivoices.com).

The study investigated the methodological implications of design interventions in community contexts, and assessed in particular: (1) the conditions, patterns, and impacts of community participation in design; and (2) the challenges of designing culturally representative digital artefacts. Drawing on Manovich [11], *culturally representative* refers to the quality of a digital artefact or system to properly reflect local ways of knowing and being, and construct in the mind of the onlooker the image of a referent existing in reality, in this case a community's socio-cultural and historical features. In an expanded sense, *culturally representative* also stands for the capacity of a system or artefact to advance locally defined goals for information, expression and communication.

To investigate these aspects, we surveyed the more generic PD literature, as well as the narrower body of research focusing on community-based participatory design. The PD literature contains a number of contributions that conceptualize the design process around the designers' and the users' inputs. Muller [1] describes PD as the third space in HCI, a space in between the developers' and the users' worlds. He makes an overview of "hybrid practices", PD techniques and methods that fall in between the users' and the developers' world. Methods such as workshop, drama, storytelling, and design games contribute to challenging assumptions and driving novel ideas emerged through dialogue and negotiation among users and developers. Kensing and Munk-Madsen's [12] model of user-developer communication in PD depicts design as a bridge across two different worlds: the users' (the workplace) and the developers' (the technological options). A third domain – the new system – is created throughout the project.

Community-based design studies are characterized, furthermore, by an increased sensitivity to contextual conditions. It has been argued that designing for local communities involves working with and accounting for two different and at times incompatible ways of knowing [2,3]. When the aim is to create culturally representative and locally useful artefacts, design relies on embracing a local viewpoint and allowing activities and design solutions to emerge from it [2]. Community-based design studies, in particular if conducted in indigenous contexts, agree that community participation is the real measure in the development of community-representative artefacts [2,3,5]. Community-centric design is not about applying the right method, it rather implies re-interpreting all articulations of the design process in the light of the specific situation [3,13]. Design in community contexts calls for the creation of a new space where design approaches and techniques are re-configured during the interaction between local members and the design team [13].

These considerations frame the position embraced in this paper, and further call for the central questions it poses: *if participatory design in community contexts is an emanation of present interactions between designers and local people, can we identify the most important factors underpinning this interaction? In particular, what are the pre-conditions for effective joint work of designers and local communities in the creation of novel artefacts?* To answer these questions, the next section introduces an analytical model of the design continuum drawing on Wenger's theory of cultural transparency [7,8], illustrated by vignettes from the Romani Voices study.

The term *community* is used in the forthcoming part with two different meanings: (1) the term "community" on its own or in the constructs "designers' community" and "users' community" reflects the meaning invested by Wenger [7,8], and indicates a group of people with common interests engaging in shared practices; (2) "Local community" is defined as a construct blending two features: shared geographical space, and members' relatedness [14], therefore referring to groups of people living in the same space and engaging in social interaction and shared activities.

3 The Theory of Cultural Transparency and PD

Étienne Wenger's theory of cultural transparency [7,8] presents interest for conceptualising PD due to the perspective it provides on the interplay between learning, participation and meaning production. The theory can be summed-up in five core statements: (1) *Knowledge* is not existing in the abstract but is constantly produced and enacted as part of the practice of communities, taken to be broadly groups of people that engage in shared activities around a profession or common interest; (2) The practices of a community result in the production of objects/representations invested with meaning (from symbols, terms and procedures to gestures and tangible artefacts); (3) *Representations* are the results of processes of reduction, simplification or meaning investment, meant to facilitate members' joint activities and interaction. They are imbued with meaning, but are at the same time hermetic for a new-comer; (4) Understanding the meaning of representations, or learning, can *only* happen through direct participation in the practices of a community; (5) *Cultural transparency* refers to the quality that objects stemming from the practice of communities acquire when their cultural significance is unveiled by an agent. It indicates the successful outcome of a process of learning, on virtue of which a community member can handle representations in a seamless manner. The term "cultural" is used to convey the context-specific and perspective-bound character of knowledge and learning [7, p. 104].

The sections below (1) outline an interpretation and operationalization of Wenger's cultural transparency theory through encoding and decoding meaning, (2) show its relevance for design, specifically PD, and (3) suggest implications for conducting effective PD processes in community contexts.

3.1 Encoding and Decoding Meaning

Central to Wenger's theory are two alternate processes that characterize learning through participation in a community of practice: (1) members' regular engagement in activities or practices relying on the production and use of representations; and (2) the new comer's attempt to learn and seize the meanings embedded in these representations. We can look at these alternate processes of *creation* vs. *understanding* of representations as a process of *encoding* vs. *decoding* meaning and significance. (Note: The encoding-decoding interpretation is part of the authors' approach and was not used by Wenger.)

1. *Encoding*: Objects and representations are created through abstraction and simplification of the practices in a given community. In this process, they are invested with significance that is invisible to an outsider's eye.
2. *Decoding*: The attempt to understand the significance of objects or representations of a community equals to a process of decoding, where an outsider tries to make out the hidden layers of meaning in a representation. This understanding is

possible, in Wenger's view, only through participation in the practices of a community, and is entirely achieved in full participation. Decoding activities build up understandings until reaching *cultural transparency*.

3.2 The Perspective on Design

Design aims to produce useful artefacts for a certain class or community of users, usefulness which can be determined only in direct relation to the user activities that the artefact will support. Effective design needs to build on a thorough understanding of the user activities that its outcome will serve. In PD, this understanding is advanced by enabling users' direct participation in design activities.

Based on the theory of cultural transparency, *participatory design* brings together the members of two communities: designers and users, each with a history of past intra-community practices and handling representations created on virtue of such practices. In instantiating PD, a new space of interaction is created, in which the practices and representations of each group are at the outset little known for the other (Fig. 1). PD activities alternate constantly between *decoding* activities, having as outcome understandings of the other community, and *encoding* activities, having as outcome new design representations. Decoding activities serve to build up to increased understanding until *cultural transparency* is reached, indicating the point where each community understands each other's representations sufficiently for engaging in fruitful design activities. The main thesis put forward in this paper is that effective design starts when cultural transparency is reached. The underpinnings and implications of this position are outlined below.

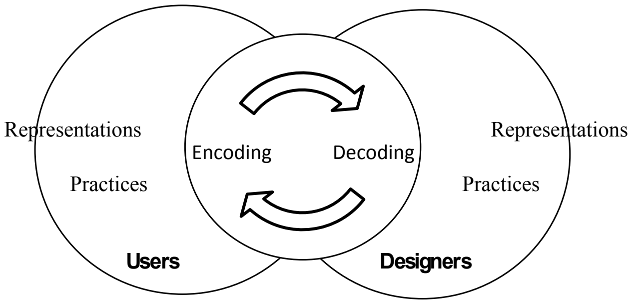


Fig. 1. Participatory design as an iterative process of encoding and decoding involving users and designers

Decoding and Encoding in the PD Process. *Decoding activities* can be categorized with respect to two dimensions: the *agent*, or who engages in learning (users or designers); and in relation to which *area of practice* (users' or designers' community). There result four types of activities: (1) Users understanding design (e.g. through demos, informative sessions); (2) Users deriving new understandings of their own context (e.g. through active questioning and reflective sessions); (3) Designers understanding the local context (e.g. through ethnographic methods); (4) Designers deriving new understandings of design in the light of the local context (e.g. through discussions on contextual applications of design solutions).

Encoding activities serve to create new representations as part of the design process. While Wenger uses the terms "objects" and "representations" interchangeably to encompass all classes of objects resulting from practice, we propose that for being useful in analysing design processes, two distinctions are necessary:

1. Based on level of concreteness, representations can be abstract (e.g. design terminology, work procedures) or concrete (e.g. written codes, manuals).
2. Based on role in the PD process, there can be *mediating* and *final representations. Mediating representations* act as intermediary in the relationship between agents (such as a procedure telling agents what each should do) or in the activity performed by agents (as in the use of a mock-up for advancing understanding of the functionalities of the final system). *Final representations* are those that users will actually be utilizing, such as the final system (Table 1).

Table 1. Examples of types of representations produced during PD, according to level of abstractness (abstract – concrete) and role in the PD process (mediating – final)

	Mediating	**Final**
Abstract	Design terms, work procedures	Workflows based on product usage
Concrete	Mock-up	Final system

Mixed activities, blending decoding and encoding can be employed in order to generate experiential knowledge and showcase or probe preliminary outcomes of design activities. Examples are cultural probes [15], in which users employ low or digital technology to document their context and produce representations that can trigger group discussions and reflections. In *Romani Voices*, local people used video technology to record snapshots of their life and traditions that were therefore discussed with the field researcher. This served a three fold purpose: (1) local people generated new understandings of their context; (2) the field researcher derived understanding of the local context through people's perspectives; and (3) local people got acquainted with the use of digital technology. Another example is the creation of a community blog and the website mock-up, both of which served to enhance understanding of design-specific terms and artefacts for local people.

A further delineation of activity types can be done by specifying the dynamics of interaction between users and designers in performing either decoding or encoding activities. These activities can be performed separately by each agent (e.g. a designer learning about the local context through non-participant observation) or jointly (e.g. the deployment and joint assessment of an information system). The latter can be referred to by Wenger's concept of *boundary practice,* indicating the sphere of interaction created through the involvement of users and designers in common activities. A boundary practice is at the same time (1) a fertile field for *enhancing understanding* of representations commonly grounded in the domain of each community; and (2) the field of *production of new meanings*. Just as any practice of a community, it can generate its own representations, understood jointly by designers and users. PD activities performed by users, designers, or jointly move constantly between the distinct areas of practice of the two communities and the boundary practice area.

The Limits of Participation. The forms that participation in design can take can be related to Wenger's concept of *legitimate peripheral participation* – LPP [7,8]. LPP as an analytical category refers to learning through increased participation in the practices of a community. However, an agent may participate in practices without attempting to become a member. Her/his position is clearly outlined and so are the limits to her/his participation, based on agreed protocols with the users. In PD, the limits of participation are defined in relation to the degree of knowledge needed by users and designers in order to cooperate effectively in the achievement of the design goals.

3.3 Implications for Design with Local Communities

The main argument put forward in this paper is that *cultural transparency* is the landmark for effective design. In PD, *cultural transparency* indicates the outcome of a process of learning focused on acquiring knowledge circumscribed to two spheres – the users' context and ICT design. Users will be able to take informed decisions regarding the new system only after having seized the meanings and significance of design representations (terminology, procedures, rules, artefacts). Designers, on the other hand, will be able to put forward adequate solutions when they have reached an understanding of users' practices.

Cultural transparency as landmark for effective design becomes critical when PD is employed in community contexts, aiming to produce artefacts that can advance long-term community goals. When design artefacts are intended to serve cultural representation as defined earlier, there is a need to draw on a common pool of understandings enabling relations between community practices and design solutions. The analytical model introduced drives attention to three aspects to be considered:

1. *The types of activities employed in PD*, which may fall in three categories: decoding, encoding, and mixed. We argue in particular for the usefulness of mixed activities (e.g. cultural probes), in which mediating representations are created for advancing experiential knowledge and understanding the applicability of design solutions to a specific context.
2. *The sequence of decoding, encoding, and mixed activities.* Effective sequencing requires sensitivity and attention to the type of knowledge and understandings required for taking informed decisions in design. We suggest that in the beginning of a project cyclic iterations of decoding and encoding activities concerned with the production of mediating representations are effective drivers of understanding on both sides. On this basis, encoding-intensive activities may further be devised.
3. *The value of boundary practices.* When engaging in joint activities, in particular those blending decoding and encoding (e.g. engaging in the deployment and discussion of cultural probes), users and designers instantiate shared spaces of interaction, with their own production of meaning. These spaces are particularly fruitful for advancing understandings of design, the local context, and design solutions grounded in the needs of the local context.

4 Conclusion

This paper introduced an analytical model for PD processes, drawing on Wenger's theory of cultural transparency. The model depicts the PD continuum as an alternation of activities in which both users and designers engage, respectively for (1) deriving knowledge and understanding of the other's area of practice and (2) investing meaning in the creation of new representations. The model is proposed as an analytical tool for understanding the factors that are brought to bear and conditions to be taken into account when engaging in PD projects in new contexts, for instance with local communities. The importance of reaching a shared sphere of understandings, indicated by the term *cultural transparency,* is especially critical when design interventions are aimed to produce culturally representative artefacts that will help to advance a community's goals and can be integrated and used in its practices.

References

1. Muller, M.J.: Participatory design: The third space in HCI. In: The Human Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications, pp. 1051–1068 (2002)
2. Rodil, K., Winschiers-Theophilus, H., Jensen, K.L.: Enhancing cross- cultural participation through creative visual exploration. In: Procs. PDC 2012, pp. 81–90. ACM (2012)
3. Verran, H., Christie, M., Anbins-King, B., van Weeren, T., Yunupingu, W.: Designing Digital Knowledge Management Tools with Aboriginal Australians. Performative Knowledge Making. Charles Darwin University, Australia (2006)
4. Christen, K.: Archival Challenges and Digital Solutions in Aboriginal Australia. The SAAA Archeological Record. Special Issue International Cooperative Research 8(2), 21–24 (2008)
5. Srinivasan, R.: Where Information Society and Community Voice Intersect. The Information Society 22(5) (2005)
6. Winschiers, H.: The challenges of participatory design in an intercultural context: Designing for Usability in Namibia. In: Proc. PDC 2006 (2006)
7. Wenger, E.: Towards a Theory of Cultural Transparency. Unpublished Doctoral Dissertation. Department of Sociology. University of California, Irvine (1991)
8. Wenger, E.: Communities of Practice: Learning, Meaning, and Identity. Cambridge University Press, Cambridge (1998)
9. Sabiescu, A.G.: Exploiting the Intergenerational Connection in Community Media Initiatives for Minority Cultures: A Case Study. International Journal of Technology, Knowledge and Society 7(6), 1–18 (2012)
10. Sabiescu, A.G.: The Issue of Collective Agency in Community-based Open Content Creation. In: Proc. IDIA 2011 (2011)
11. Manovich, L.: The language of new media. MIT Press, Cambridge (2001)
12. Kensing, F., Munk-Madsen, A.: PD: Structure in the toolbox. Com. of the ACM 36(6), 78–85 (1993)
13. Van der Velden, M.: Design for the contact zone: Knowledge management software and the structure of indigenous knowledges. In: Proc. CATaC 2010, pp. 1–18 (2010)
14. Gusfield, J.R.: The community: A critical response. Harper Colophone, New York (1975)
15. Gaver, B., Dunne, T., Pacenti, E.: Cultural Probes. Interactions 6(1), 21–29 (1999)

Reducing Driver Task Load and Promoting Sociability through an Affective Intelligent Driving Agent (AIDA)

Kenton Williams and Cynthia Breazeal

Massachusetts Institute of Technology, MIT Media Lab, 20 Ames St., E15-468, Cambridge, MA, United States of America
{kjwillia,cynthiab}@media.mit.edu

Abstract. This work outlines the development of an Affective Intelligent Driving Agent (AIDA), a social robot that sits in a vehicle's dashboard and behaves as a friendly assistant. This highly expressive robot uses an Android smartphone as its face, which serves as the main computational unit for the system. AIDA determines what information may be relevant to the driver, delivers it at the most appropriate time, and resolves which expressions should be used when doing so. An evaluation was performed in which participants completed mock driving tasks with the aid of 1) a smartphone with apps, 2) AIDA as a static, expressive agent, or 3) AIDA as a mobile robot. Results showed that the AIDA robot helped reduce user task load and promoted more sociability with users better than the smartphone or AIDA as a static agent.

Keywords: Affective HCI, Human-robot interaction, Novel user interfaces and interaction techniques.

1 Introduction

Given the substantial amount of time people spend driving, there is an increasing demand for enjoyable in-car experiences that do not compromise safety. To make better use of their time, drivers tend to multi-task (e.g. checking traffic conditions, exchanging texts, etc.) by manipulating their In-Vehicle Infotainment (IVI) systems or mobile devices. Interacting with such devices while driving results in high cognitive load, which can amplify the driver's lack of focus and could lead to accidents.

While several IVI systems have been designed to keep the driving experience as safe as possible, it is still inconvenient for drivers to limit themselves to the systems that are part of the car. Many users would like to access their favorite applications everywhere, particularly inside their cars, despite the fact that some of these applications were not designed for a driving context. Many vehicles fail to provide a seamless transition for drivers as they enter and exit the vehicle since data usually does not flow into the car unless additional devices are brought in, and used while driving.

In addition to safety, social and emotional factors play a vital role in the driving experience, yet many IVI systems neglect these influences. For instance, car manufacturers do consider social factors of their driver demographic, like personality and

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 619–626, 2013.

emotion, as key aspects of car design [1]. Further, the driver's emotional state can have a significant impact on driver behavior and safety. Tired or stressed drivers are less likely to fully focus on the road, as compared to alert drivers. And angry, frustrated drivers are more prone to road rage, as compared to happy drivers.

These issues indicate a need for an IVI system that reinforces safety, manages information from the environment and the driver's personal device, and interacts with the driver in a social manner. Integrating these aspects into a unified architecture could make the driving experience safer, more efficient, and overall, more enjoyable.

2 Related Work

Tchankue, Wesson and Vogts addressed driver distraction while texting by presenting information upon arrival, and additional cues were included to express the priority level of the message [2]. These cues would allow the driver to decide when to deal with the received message. They found that these informative interruption cues were learned quickly and identified accurately to reduce cognitive load. However, this work does not explore the social aspect of driving and the possibility of taking proactive actions that reach out to other applications to offer more support to the driver.

Nass and Brave concluded that matching driver emotions with the IVI system's tone of voice improves overall in-car performance [3]. Matching happy drivers with an enthused voice and upset drivers with a subdued voice resulted in superior driving performance as opposed to matching opposite pairs. In [4], affective speech interfaces were explored as a means of seamless driver-vehicle communication. While these works reveal the benefits of speech-based interfaces, they do not explore information delivery in an expressive, sociable way to improve the quality of communication. If the IVI can express itself in a way that feels natural and familiar to the driver, then there could be a deeper understanding of the messages that it conveys.

Technologies like Bluetooth headsets and docking stations on windshields and dashboards allow drivers to access their mobile devices without compromising safety [5]. In [6], a system was designed that integrates the IVI system and the user's smartphone. Mobile devices run all phone services, while the IVI system is responsible for all input/output functions to handle a wider range of applications. While these systems are convenient, they degrade the user experience because they are highly reactive. The driver still has to initiate many actions to obtain necessary information.

3 Core Contribution

With motivations in mind from the previous sections, we have developed an Affective Intelligent Driving Agent (AIDA) [7], a socially expressive robot that acts as a friendly, in-car assistant, Fig. 1. AIDA uses the driver's mobile device as its face. Thus, AIDA uses app-based functionality to display facial expressions, manage and deliver task-specific information, and leverage aspects of the driver's daily life outside of the vehicle for deep personalization.

Fig. 1. In-car concept of the Affective Intelligent Driving Agent (AIDA)

AIDA allows users to accomplish their driving tasks while promoting in-car safety by keeping the phone out of the driver's hands. Since AIDA acts as an intermediary between the driver, the phone's applications, the vehicle and the environment, people will not lose access to the applications they need to perform their activities. Further, as opposed to traditional methods of driver-vehicle communication, e.g. flashing icons and chiming sounds, AIDA communicates with the driver through speech, coupled with expressive body movements.

4 System Framework

The robot consists of a five degree-of-freedom head and neck mechanism, which sits seamlessly in the dashboard when recessed, but conveys stronger expressions when extended. An Android smartphone acts as the face of the robot. The robot's plastic head shell was designed so that the driver can easily snap in the mobile device.

The Android phone is the core of AIDA's framework, which manages internal apps and sensor data, as well as communicates with the physical robot. We have developed an AIDA app, which encapsulates all of this functionality, displays facial expressions, and interconnects other apps. The AIDA app also queries external sources like the Internet for further assistance. The driver communicates with the device via speech commands and to a lesser extent, through tactile interactions.

In addition to the phone's standard apps, we have developed AIDA-specific apps (excluding the LTTS Susan app which was purchased through the Google Play Store) to assess AIDA's effectiveness as a proactive, social driving companion.

(a) LTTS Susan **–** A text-to-speech engine used to convert written text to verbal utterances using a female voice. This app is used to verbalize emoticons and text statements that the driver may receive during the driving task.
(b) Gas Station Finder (GSF) – Finds nearest gas stations given the driver's current location. Once the driver selects a station, the app navigates them appropriately.
(c) Travel Time Calculator (TTC) – Determines how long it will take for the driver

to navigate to a certain location. The driver inputs an address and the app determines their travel time given the driver's current GPS coordinates.

(d) AIDA –This app integrates all context-specific apps into a seamless architect ture, while an expressive face acts as the main interface. The driver interacts with this app via speech and to a lesser extent, through touch.

(e) R1D1 – A communication hub between the phone and robot [8]. It sends motor positions to the robot given commands from the AIDA app.

5 Experiment

Our experiment examines peoples' attitudes towards AIDA and assesses user interactions with AIDA compared to a mobile device during mock driving tasks. For now, we are not assessing AIDA's impact on driver car handling (e.g. steering, response time, etc.). Instead, we are evaluating AIDA's ability to provide a more seamless way for the driver to handle the data flow between themselves, the car, and the outside world while promoting sociability.

5.1 Protocol

Upon arrival, participants were given an orientation, which explained the procedures of the experiment. We demonstrated how to use the smartphone including all necessary apps and services. Users were allowed to practice with the phone until they felt comfortable using it for the experiment. Users were then placed in an in-car environment where a laptop mounted atop a dashboard/steering wheel rig displayed a mock driving system. Users were videotaped while they performed the 15-minute task. Last, participants filled out a questionnaire regarding the interaction.

5.2 Experimental Conditions

Three conditions were used for the evaluation, Fig. 2. In **Case PHONE**, users completed the task with the smartphone alone, Fig. 2 (left). People used the Messaging, Calendar, Contacts, GSF, and TTC apps to assist them. This is similar to the way drivers currently use their smartphones in their vehicles. In **Case AGENT**, users completed the task with the AIDA app, while the phone was mounted on a stationary

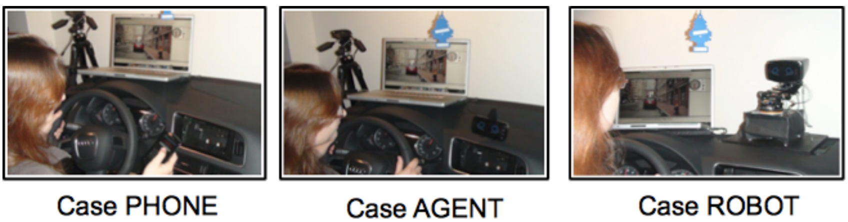


Fig. 2. Three experimental conditions

dock, Fig. 2 (middle). This resembles existing in-car embedded screens and docking stations for driver assistance. In **Case ROBOT**, users completed the task with the AIDA app attached to the physical robot, Fig. 2 (right). To minimize distractions, the robot was programmed to move only when the car was stationary in the mock driving system. This explores an in-car interface that fuses driving assistance with sociability.

5.3 Experimental Task

Users had two events to attend while stress-inducing phone alerts and vehicle warnings occurred throughout the interaction. We encouraged users to act as naturally as possible and to perform any in-car actions (i.e., signaling, lane switching, accelerating, braking, etc.) that they would normally do while driving in the real world. Thus, users had the choice to make/answer phone calls or to send/read text messages. The R1D1 app handled communication between the driving system and phone/AIDA.

Two events were pre-programmed into the phone's calendar: 1) Lunch with the user's fictitious friend named Jaylen and 2) Salsa dancing lessons with another fictitious friend named Nancy. The locations and contact information associated with these events were also stored. The participant's goal was to arrive at these two events on time in the mock driving system. However, we crafted the system to force the user to become late for their first event due to traffic, irrespective of their driving performance. We also fashioned a low-gas vehicle warning when the user navigated to the second event. We also made phone calls and sent text messages to the user (pretending to be Nancy and Jaylen) at critical moments throughout the task. We monitored the interaction remotely in order to send these alerts at the appropriate times.

In **PHONE**, to navigate to the respective events, participants launched the Calendar app to find the appropriate event addresses. Once they knew the address, they then launched the TTC app to figure out their estimated travel time. We programmed the mock driving system to activate only once the user's travel time was established. If users chose to respond to our texts/calls throughout the interaction, they had to do so while trying to focus on the driving task. If users wanted to fill their gas tank in response to the low-gas warning, they launched the GSF app to assist them.

In **AGENT**, participants launched the AIDA app and AIDA proactively navigated the user to the event, which made beginning the task more seamless. Because the AIDA app encapsulates the Calendar and TTC apps, AIDA is aware of the driver's contacts and calendar events before they even enter the vehicle. Further, contextual awareness allows AIDA to better mitigate data flow between the driver and outside environment. For example, when the driving system notified the driver that they were late for their first event, AIDA preemptively asked the user if they wanted to send a text to Jaylen (i.e. us) notifying him of their tardiness. This way, users could communicate with the outside world while focusing on the driving task. Also, if users wanted to fill their tank in response to the low-gas warning, AIDA preemptively navigated them to the nearest station since the GSF app is integrated into the AIDA app.

In **ROBOT**, the interaction was similar, as previously described in **AGENT**. However, in addition to proactive assistance, the AIDA robot elicited direct non-verbal cues to the driver via body movements. Expressive body movements were coupled with AIDA's messages to the driver to promote deeper sociability.

5.4 Hypotheses

Taking insights from works in anthropomorphic interface technology, we predict three hypotheses. In [9], it was determined that affective interfaces had a major impact on reducing mental overload. Relating this work to our driving task, we predict:

H1 – Task Load: AGENT and ROBOT users will find the experimental task less mentally demanding and easier to complete than PHONE users.

In [10], people felt a stronger sense of cooperation and sociability with a robotic assistant versus an expressive agent. Thus, we predict:

H2 – Co-Supportive Teaming: AGENT and ROBOT users will see AIDA as an equal partner more than PHONE users with the smartphone. Further, this sentiment will be stronger felt in ROBOT than in AGENT.

H3 – Impact of Embodiment: ROBOT users will feel a deeper social bond with AIDA as a robot than AGENT users with AIDA as an on-screen agent only.

5.5 Dependent Measures

Hypotheses were evaluated through questionnaire responses and video footage. Survey questions from accepted sources provided us with standard methods to assess sociability and cooperation [11] and cognitive load [12]. Video data was used to assess user attention and mood. We coded for eye gaze, facial expressions, gestures, utterances and in-car actions. We define positive affect as smiles, laughs, excitement and positive utterances/gestures expressed by participants. Conversely, we define negative affect as frowns, confusion, anger, sadness and negative utterances/gestures.

6 Results

We recruited 44 participants (13 PHONE, 17 AGENT, and 14 ROBOT) from the {location} area. There were 20 males, 24 females and the mean age was 28.6 years.

Questionnaires were scored on a 5-point Likert Scale (1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree). Mean (*M*) and Standard Deviation (σ) values of user responses were calculated for each experimental condition, Fig. 3. ANOVA *F*-values (AFV) were calculated to assess global variance. Pair-wise comparisons were then found using the Tukey-Kramer Method (TKM).

Video footage was transcribed by two objective individuals for behaviors discussed in Section 7. Krippendorff's alpha criterion was used to determine inter-coder reliability where α-values between 0.8 and 1 (inclusive) suggest acceptable consistency. We found $\alpha = 0.873$ for key behaviors highlighted in Table 1. Pair-wise comparisons between PHONE-AGENT (T_{PA}), PHONE-ROBOT (T_{PR}) and AGENT- ROBOT (T_{AR}) were then found using TKM, Table 1.

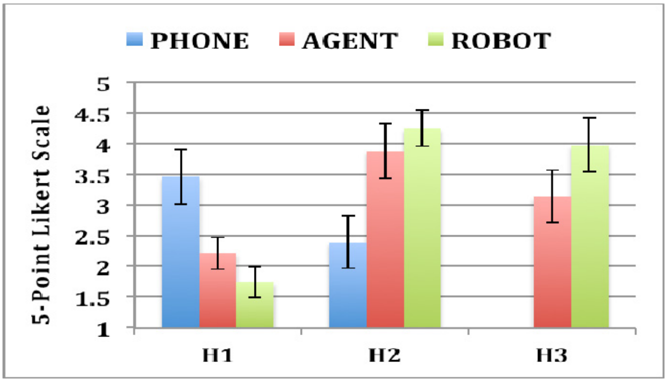


Fig. 3. Mean (*M*) and Standard Deviation (σ) of questionnaire responses (error = $\pm\ \sigma$)

Table 1. Observations from video data with Tukey-Kramer Method (TKM) comparisons

Behavior	PHONE	AGENT	ROBOT	T_{PA}	T_{PR}	T_{AR}
a) Positive Affect	7.0	13.5	16.5	6.5	9.5	3.0
b) Negative Affect	5.5	2.0	1.5	3.5	4.0	0.5

H1 - Fig. 3 (H1) shows the average replies to questions regarding high task load during the experiment. AFV revealed that most PHONE users felt that they experienced a high task load while AGENT and ROBOT users largely disagreed ($F(2,41) = 3.88$, $p = 0.029$). TKM comparisons showed significant differences between each pair-wise group (T(40,3), all $p < 0.05$). Thus, ROBOT users felt less exhausted by the task than other users. Table 1 (b) illustrates that PHONE users expressed more negative affect throughout the task. TKM comparisons, T(40,3), showed significant differences between PHONE-AGENT and PHONE-ROBOT, but not AGENT-ROBOT.

H2 - Fig. 3 (H2) shows the mean responses to questions regarding the user's feeling that they were working as a team with the phone or AIDA to accomplish the driving task. PHONE users did not feel a sense of equal partnership with the phone as compared to AIDA users ($F(2,41) = 10.26$, $p = 0.0003$). TKM comparisons, T(40,3), showed significant differences between all pair-wise groups. Thus, users felt that the AIDA robot was more of a supportive teammate than the static AIDA display.

H3 - Table 1 (a) shows that AIDA users expressed more positive affect as compared to PHONE users ($F(2,41) = 4.92$, $p = 0.039$). TKM comparisons, T(40,3), revealed that ROBOT users expressed significantly more positive affect than AGENT users. Fig. 3 (H3) shows the mean responses to questions regarding the feeling that a mutual social bond existed between the user and AIDA. Users felt more mutual social awareness with the AIDA robot than the static display ($F(1,29) = 5.89$, $p = 0.003$).

7 Discussion

H1 – Task Load: Validated. Since AIDA users did not need to manipulate the device as much as PHONE users, they found the task less mentally demanding.

H2 – Co-Supportive Teaming: Validated. Due to its proactive support, AIDA was seen as an equal partner during the mock driving task more than the smartphone. This sentiment was stronger felt with the AIDA robot than the static agent.

H3 – Impact of Embodiment: Validated. The AIDA robot was able to elicit more personally directed non-verbal cues to the driver. Thus, users felt a deeper social bond with the embodied robot than the static display.

This work provides initial insights into AIDA's effectiveness at reducing task load and promoting sociability. Our results suggest that AIDA has the potential to better support drivers than smartphones alone during simulated driving tasks. They also suggest that the AIDA robot can promote sociability better than the static agent. Further investigation will explore AIDA's effectiveness during real-world driving tasks.

References

1. Bayley, M., Curtis, B., Lupton, K., Wright, C.C.: Vehicle Aesthetics and their Impact on the Pedestrian Environment. Technical report, Transportation Research D (2004)
2. Tchankue, P., Wesson, J., Vogts, D.: The Impact of an Adaptive User Interface on Reducing Driver Distraction. In: International Conference on Automotive User Interfaces, Salzburg, Austria (2011)
3. Nass, C., Brave, S.: Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship. MIT Press, Michigan (2005)
4. McAllister, G., McKenna, S.J., Ricketts, I.W.: Towards a Non-Contact Driver-Vehicle Interface. In: Intelligent Transportation Systems (2000)
5. Bühler, D., Vignier, S., Heisterkamp, P., Minker, W.: Safety and Operating Issues for Mobile Human-Machine Interfaces. In: International Conference on Intelligent User Interfaces, Miami, Florida (2003)
6. Bose, R., Brakensiek, J., Park, K.Y.: Terminal Mode – Transforming mobile devices into automotive application platforms. In: International Conference on Automotive User Interfaces, Pittsburg, Pennsylvania (2010)
7. Foen, N.: Exploring the Human-Car Bond through an Affective Intelligent Driving Agent. Master's thesis. MIT, Cambridge, MA (2012)
8. Isla, D., Burke, R., Downie, M., Blumberg, B.: A layered brain architecture for synthetic characters. In: The International Joint Conference of Artificial Intelligence (2001)
9. Nass, C., Steuer, J., Tauber, E.R.: Computers are social actors. In: Special Interest Group on Computer Human Interaction: Human Factors in Computing (1994)
10. Kidd, C., Brezeal, C.: Comparison of social presence in robots and animated characters. Interaction Journal Studies (2005)
11. Takayama, L., Groom, V., Nass, C.: I'm sorry, Dave: I'm afraid I won't do that: social aspects of human-agent conflict. In: Human Factors in Computing Systems. ACM, New York (2009)
12. Sweller, J.: Cognitive load during problem solving: Effects on learning. In: Cognitive Science, vol. 257 (1988)

Semantic Modelling in Support of Adaptive Multimodal Interface Design

Elena Tsiporkova[1], Anna Hristoskova[2], Tom Tourwé[1], and Tom Stevens[3]

[1] Sirris – Software Engineering & ICT Group
A. Reyerslaan 80 – 1030 Brussels – Belgium
{elena.tsiporkova,tom.tourwe}@sirris.be
[2] Department of Information Technology, Ghent University – iMinds
Gaston Crommenlaan 8 (201) – 9050 Ghent – Belgium
anna.hristoskova@intec.UGent.be
[3] Namahn – Agency for Human-Centered Design
Grensstraat 21 – 1210 Brussels – Belgium
ts@namahn.com

Abstract. The design of multimodal interfaces requires intelligent data interpretation in order to guarantee seamless adaptation to the user's needs and context. HMI (human-machine interaction) design accommodates varying forms of interaction patterns, depending on what is most appropriate for a particular user at a particular time. These design patterns are a powerful means of documenting reusable design know-how. The semantic modelling framework in this paper captures the available domain knowledge in the field of multimodal interface design and supports adaptive HMIs. A collection of multimodal design patterns is constructed from a diversity of real-world applications and organized into a meaningful repository. This enables a uniform and unambiguous description easing their identification, comprehensibility and applicability.

Keywords: Human-machine interface, Multimodal Design Patterns, Adaptive Interfaces, Pro-active Interaction, Data Modelling, Context-awareness.

1 Introduction

The design of *multimodal* and *adaptive* interfaces for complex real-time applications requires a specific approach in order to guarantee that the interaction between human and computer remains natural [1]. In order for the interface to adapt to the user and the context, the system needs to reason about his/her needs and proactively adapt to these while keeping the user in control. Thus, the HMI (human-machine interaction) design needs to accommodate varying forms of interaction, depending on what is most appropriate for that particular user at that particular time.

The contribution described in this paper has been developed in the context of ASTUTE[1], a large EU research project. The project focuses on the design of intelligent multimodal user interfaces, providing pro-active decision support to the

[1] ASTUTE Pro-active decision support for data-intensive environments; http://astute-project.eu/

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 627–634, 2013.

user. The goal is to develop a platform for building embedded products that capture, reason and act upon user intentions thereby taking into account his/her context (i.e. environment and all the factors which influence his/her performance) and state (i.e. aspects determining his/her ability to perform in a given situation, such as stress level, fatigue). The project approach is validated by various industrial demonstrators in the domains of automotive, avionics, emergency dispatching, building management and manufacturing process management. One of the goals of the project is to provide an appropriate methodology and tools for user interface design, based on design patterns.

HMI design patterns play a major role in exploring and specifying the interactions between users and devices by inspiring design and enabling the reuse of concrete solutions through formal descriptions [2]. An abundance of resources collect design patterns such as Yahoo! Design Pattern Library [3], Patterny [4], search patterns [5], gesture interfaces [6], rich interaction design principles on the Web [7, 8], patterns for effective interaction design [9]. The research toward the establishment of formal principles and guidelines for multimodal interaction design is also gaining increasing interest and importance in recent years (e.g. [10-15]). However, as observed by Sarter [16], the existing guidelines mostly focus on high-level design objectives and do not provide support on how to map them to the needs of an actual application.

In order to facilitate the design of multimodal interfaces in practice, this article pursues the definition of a semantic modelling framework, which has the capacity to: 1) capture and model design patterns[2], formal guidelines and expert domain knowledge in the field of multimodal interface design; 2) reason and derive explicit recommendations in support of both the HMI design and the dynamic HMI adaptation at runtime. A collection of multimodal design patterns exploring diversity of real-world applications is constructed and organized into a meaningful pattern repository. This repository is complemented and extended with ontologies defining relevant use case specific knowledge. Two application examples of the described patterns for the case of emergency dispatching and manufacturing process management are discussed.

2 Semantic Modelling Framework for Design Patterns

2.1 Design Pattern Parameterisation

A substantial part of the work has been devoted to the development of a formal parameterisation framework allowing for the construction of an HMI design pattern repository. A set of parameters for HMI pattern description and specification has been derived through multiple interactive and iterative sessions between a mixed team of HMI designers and ontology engineers. The goal was to develop a uniform HMI pattern model reflecting the need for a formalized description and an increased level of abstraction detail. This supports the decision of pattern applicability in a certain design context. In the spirit of ontology-based modelling, the resulting model is a hierarchical class structure, describing pattern parameters and their attributes [18]:

- *Pattern Parameters*: User (Junior, Senior), Environment, Device, Task (Primary, Secondary), Interaction, Information, Modality (Input, Output), Event, System.

[2] While respecting the description structure as proposed by the Gang of Four in [17].

- *Parameter Attributes*: User Attributes (Expertise, Role, State), Environment Attributes (Safety Level, Change Rate, Noise Level), Interaction Attributes (Interaction Pattern, Interaction Type), Device Attributes (Component, Size).

Various relationships have been derived between the parameter/attribute classes and subclasses as demonstrated later in Section 3.3.

2.2 Layers of Modelling Abstraction

The proposed hierarchical parameter model and the relationships defined between the different parameter classes have been modelled as an ontology enabling further model refinement and reasoning. The method structures the ontological model in a nested set of models consisting of three levels of abstraction as presented in Fig. 1:

- **Core Design Pattern Model**: Complex ontological model containing the generic key concepts and knowledge relevant to design patterns which creation and maintenance requires solid expertise in ontology modelling. HMI designers are closely involved as a source of domain knowledge and for validation purposes.
- **Repository of Design Patterns**: A collection of relatively simple ontological models, defining the specific properties and applicability of each pattern. The main challenge lies in mapping the natural language description of a pattern into a formal logic representation without losing too much expressiveness. This requires close collaboration between ontology experts and HMI designers.
- **Use Case Specific Models:** The purpose of this level is twofold: 1) to derive design recommendations (e.g. applicable design patterns) during the interface design; 2) to allow dynamic interface adaptation (e.g. context-aware output modality) at runtime. The concrete use case is described by making use of the concepts and relationships defined in the design model. HMI designers can execute this autonomously if a suitable (e.g. web-based) interface is provided.

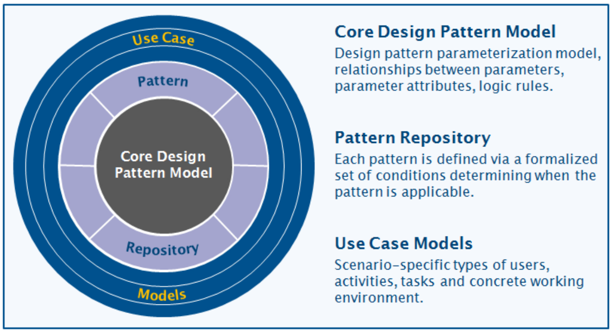


Fig. 1. Schematic overview of the different levels of semantic modelling abstraction

The core design pattern ontology model consists of 2 types of high-level key domain concepts *Parameter* and *ParameterAttribute*, describing the design pattern parameters and their attributes. These are related via the hasAttribute relationship. Each parameter is connected with its attributes through a set of relationships, which are sub-properties of this property. Some concrete examples are given in Table 1.

Table 1. Relation between specific *Parameters* and their *Attributes*

Parameter	**Relation**	**Attribute**	**Description**
User	hasExpertise	*Expertise*	expertise level of the application user, e.g. *first_time_user* or *advanced*
	hasAttention	*Attention*	focus of the user's attention, e.g. *environment, on_single_device*, *distributed_among_devices*
	canInteractVia	*InteractionChannel*	interaction capacity of the user, e.g. *hearing*, *voice*, *touch*, *sight*
Device	hasComponent	*Component*	device components e.g. *Video* (*camera*, *display*), *Haptic* (*vibration_device*)
Task	hasStructure	*Structure*	user task consists of a *single_action* or a *composition_of_actions*
Interaction	exhibitsPattern	*InteractionPattern*	applicable design pattern for the concrete situation, e.g. *important_message_pattern*, *adapting_context_detail_pattern, combination _of_modalities_pattern* (detailed in Section 5)
Environment	hasChangeRate	*ChangeRate*	frequency of change occurring in the environment, e.g. *frequent*, *never*, *rare*
	hasSafetyLevel	*SafetyLevel*	safety level of the environment, e.g. *potential_risk*, *risk_full*

The core ontology also defines a multitude of relationships between the different *Parameter* subclasses. Several concrete examples are visualized in Fig. 2.

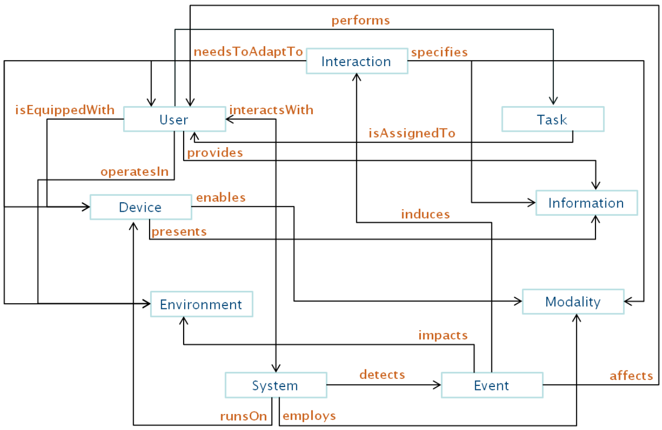


Fig. 2. Some of the relationships between the subclasses of the *Parameter* class

In addition to the class hierarchy and relationships defined between the different subclasses of the hierarchy, the core ontology is also equipped with some reasoning capacity. Two types of reasoning rules are implemented:

- Chain rules allowing to deduce a relationship based on the transitive application of two other relationships, e.g.: *(Interaction* needsToAdaptTo *User)* **AND** *(User* isEquippedWith *Device)* => (*Interaction* needsToAdaptTo *Device)*
- SWRL rules expressing some common sense knowledge or logic rules applicable to any Design Pattern, e.g.: **IF** (*System* detects *Event*) **AND** (*Event* hasCriticallityLevel *severe*) **THEN** (*Event* hasPriority *high*)

The original core ontology is complemented and extended by creating a separate ontology for each design pattern. Such an ontology needs to 1) import the Core Design Pattern Model ontology; 2) instantiate the *InteractionPattern* class with the concrete Design Pattern; 3) describe in a formal fashion (using SWRL rules) the conditions under which the concrete Design Pattern is applicable, e.g. Table 2:

Table 2. Examples of SWRL description of the applicability conditions for design patterns

Patterns	Applicability Conditions
Important Message	**IF** (*System* detects *Event*) **AND** (*Event* hasPriority *high*) **AND** (*User* hasAttention *on_environment*) **AND** (*Event* induces *Interaction*) **THEN** (*Interaction* exhibitsPattern *important_message_pattern*)
Combination of Modalities	**IF** (*System* detects *Event*) **AND** (*Event* impacts *Environment*) **AND** (*Environment* hasSafetyLevel *risk_full*) **AND** (*System* sends *Information*) **AND** (*Information* hasType *instruction*) **AND** (*Event* induces *Interaction*) **THEN** (*Interaction* exhibitsPattern *combination_of_modalities_pattern*)
Adapting Context Detail	**IF** (*System* detects *Event*) **AND** (*System* displays *Information*) **AND** (*Information* hasType *status_update*) **AND** (*User* canInteractVia *sight*) **AND** (*User* hasGoal *UserGoal*) **AND** (*Event* increasesProximityTo *UserGoal*) **AND** (*Event* induces *Interaction*) **THEN** (*Interaction* exhibitsPattern *adapting_context_detail_pattern*)

3 Use Case Models

This section provides two application examples of the described patterns for the case of emergency dispatching and manufacturing process management. The original core ontology and the repository of design patterns are complemented and extended by creating ontologies with relevant use case specific knowledge.

3.1 Emergency Dispatching Management Application

A fire commander instructs his firefighters to evacuate the building that is on fire. The ontologies discussed above are instantiated with information such as: 1) *types of users involved*: fire fighters, fire team commanders, fire station dispatchers, air sampling collectors, medical experts; 2) *activities and tasks*: evacuation, search and rescue, locating water supplies, search and rescue, defining security perimeters in the presence of dangerous substances; 3) *applications and devices:* device type (mobile, sensor), status (active, idle), components (audio, video, haptic), application features (supported users, detected events); 4) *working environment:* inside/outside, noise level, security level; 5) *events*: type (toxic smoke formation, approaching dangerous goods), priority, criticality level.

The purpose of this level is to recommend design patterns and derive new knowledge from the models and data in the different abstraction layers. For instance, an ontology describing the emergency dispatching case may contain the following:

Parameter	**Relation**	**Attribute**
fire_commander	performs	*coordination_of_evacuation*
	operatesIn	*site_on_fire*
	isEquippedWith	*fire_commander_tablet*
coordination_of_evacuation	isFocusedOn	*site_on_fire*
	isPartOf	*coordination_of_emergency*
fire_commander_tablet	hasComponent	*loudspeaker, touch_screen*
	hasStatus	*active*
	isComplementedBy	*fire_commander_smart_phone*
toxic_smoke_formation_app	detects	*toxic_smoke_formation_event*
toxic_smoke_formation_event	hasCriticalityLevel	*severe*
	impacts	*site_on_fire*
	affects	*fire_commander*
	induces	*toxic_smoke_formation_interaction*
SWRL rules	**IF** *(System detects toxic_smoke_formation_event)* **THEN** (*System* sends *toxic_smoke_formation_message*)	

Then the following additional information and recommendations will be derived:

Parameter	**Relation**	**Attribute**
toxic_smoke_formation_app	interactsWith	*fire_commander*
	informs	*fire_commander*
toxic_smoke_formation_event	hasPriority	*high*
toxic_smoke_formation_interaction	exhibitsPattern	*important_message_pattern*
	specifies	*body_awareness, visual_out*
toxic_smoke_formation_message	isPresentedOn	*fire_commander_tablet*
	hasType	*alarm*

3.2 Manufacturing Process Management Application

A team leader monitors a production line. By approaching it the system zooms in and offers richer information about its productivity, such as the Overall Equipment Efficiency (OEE), which defines the quality, performance and availability of the line. The design pattern ontologies are instantiated with the following information: *1) types of users involved*: team leaders, production line workers, maintenance worker; 2) *activities and tasks*: monitoring of and working on the production line, repairing machines; 3) *applications and devices:* device type (mobile, sensor), status (active, idle), components (audio, video, haptic, display, keyboard), application features (users supported, events detected); 4) *working environment:* production line space, administration office, visibility level; 5) *events*: approaching production line.

The production management ontology contains the following data for deriving design recommendations and knowledge from the abstraction layers' models:

Parameter	Relation	Attribute
team_leader	performs	*monitors_production_line*
	isEquippedWith	*team_leader_tablet*
	canInteractVia	*sight*
monitors_production_line	isFocusedOn	*production_line*
	requires	*production_line_overview_screen*
	isAssignedTo	*team_leader*
production_line	hasVisibilityQuality	*excellent*
team_leader_tablet	hasComponent	*loudspeaker, touch_screen*
	hasStatus	*active*
approach_production_line_app	detects	*approach_production_line_event*
approach_production_line_event	hasCriticalityLevel	*moderate*
	impacts	*production_line*
	affects	*team_leader*
	induces	*approach_production_line_interaction*
	increasesProximityTo	*production_line*
production_line_overview_screen	hasType	*status_update*
	isAbout	*overall_equipment_efficiency_oee*
SWRL rules	**IF** *(System* detects *approach_production_line_event)* **THEN** *(System* sends *production_line_context_details)*	

Consequently, the following information and recommendations are derived:

Parameter	Relation	Attribute
approach_production_line_app	interactsWith	*team_leader*
	employs	*visual_out*
	runsOn	*team_leader_tablet*
	operatesIn	*production_line*
	sends	*production_line_context_details*
approach_production_line_interaction	specifies	*production_line_context_details*
	exhibitsPattern	*adapting_context_detail_pattern*
	needsToAdaptTo	*team_leader, team_leader_tablet*
production_line_overview_screen	isPresentedOn	*team_leader_tablet*
	hasDetailGranularity	*detailed*

4 Conclusion

This paper presents a semantic technology approach to formally model and exploit relevant domain knowledge about HMI design patterns. Our aim is to demonstrate that the modelling of appropriate data about user state and context linked to the specification of interaction patterns constitute a powerful means to realise a proactive, adaptive multi-modal interaction both at design time and runtime. Besides further refinement of the presented semantic modelling framework, we intend to open our pattern repository to the benefit of the HMI community and thus enable different stakeholders in the design domain to interact, exchange ideas, improve and annotate patterns proposed by others, contribute new patterns and describe other use cases.

References

1. Plass-Oude Bos, D., Poel, M., Nijholt, A.: A Study in User-Centered Design and Evaluation of Mental Tasks for BCI. In: Lee, K.-T., Tsai, W.-H., Liao, H.-Y.M., Chen, T., Hsieh, J.-W., Tseng, C.-C. (eds.) MMM 2011 Part II. LNCS, vol. 6524, pp. 122–134. Springer, Heidelberg (2011)
2. Goodwin, K.: Designing for the digital age: How to create human-centered products and services. Wiley (2011)
3. Yahoo! Design Pattern Library, `http://developer.yahoo.com/ypatterns/`
4. Oy, P.F.: Patterny, Pattern Factory Oy (Ltd.), `http://patternry.com/`
5. Morville, P., Callender, J.: Search Patterns: Design for Discovery. O'Reilly Media Inc. (2010)
6. Saffer, D.: Designing Gestural interfaces. O'Reilly Media Inc., Sebastopol (2009)
7. Scott, B., Neil, T.: Designing Web Interfaces, Principles and Patterns for Rich Interaction. O'Reilly Media Inc. (2009)
8. Welie, M.V.: Patterns in interaction design - Patterns library. (2008), `http://www.welie.com/`
9. Tidwell, J.: Designing interfaces, Patterns for effective interaction design. O'Reilly Media Inc. (2010)
10. Godet-Bar, G., Dupuy-Chessa, S., Nigay, L.: Towards a system of Patterns for the design of Multimodal Interfaces. In: Computer-Aided Design of User Interfaces V, Proceedings of the Sixth International Conference on Computer-Aided Design of User Interfaces CADUI 2006, Bucharest (2006)
11. Ratzka, A., Wolff, C.: A Pattern-Based Methodology for Multimodal Interaction Design. In: Sojka, P., Kopeček, I., Pala, K. (eds.) TSD 2006. LNCS (LNAI), vol. 4188, pp. 677–686. Springer, Heidelberg (2006)
12. Ratzka, A.: Identifying User Interface Patterns from Pertinent Multimodal Interaction Use Cases. Mensch und Computer 2008: Viel Mehr Interaktion (2008)
13. Dumas, B., Lalanne, D., Oviatt, S.: Multimodal interfaces: A survey of principles, models and frameworks. In: Lalanne, D., Kohlas, J. (eds.) Human Machine Interaction. LNCS, vol. 5440, pp. 3–26. Springer, Heidelberg (2009)
14. Oviatt, S.: Multimodal interfaces. In: The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications, pp. 286–304 (2003)
15. Reeves, L., Lai, J., Larson, J., Oviatt, S., Balaji, T., Buisine, S., Collings, P., Cohen, P., Kraal, B., Martin, J., et al.: Guidelines for multimodal user interface design. Communications of the ACM 47(1), 57–59 (2004)
16. Sarter, N.: Multimodal information presentation: Design guidance and research challenges. International Journal of Industrial Ergonomics 36(5), 439–445 (2006)
17. Gamma, E., Helm, R., Johnson, R., Vlissides, J.: Design patterns: elements of reusable object-oriented software. Addison-Wesley (1994)
18. Ferreira, N., Geldof, S., Stevens, T., Tourwé, T., Tsiporkova, E.: Patterns for HMI design of multi-modal, real-time, proactive systems. In: Proceedings of the IUI 2013 Workshop on Interacting with Smart Objects (2013)

Supporting Improved Maternity Care by Midwives: Design Opportunities and Lessons Learned

Abdullah Al Mahmud and David V. Keyson

ID-StudioLab, Industrial Design Engineering
Delft University of Technology, 2628 CE Delft, The Netherlands
{a.almahmud,d.keyson}@tudelft.nl

Abstract. In this paper we describe a study about the role of Information Technology (IT) on the quality of maternity care in the midwifery centres in the Netherlands. We conducted an interview study with midwives in the Netherlands. The objective was to understand the current situation, challenges and design opportunities that could help to provided improved healthcare. The results of the interview study show that the current prenatal care system suffer from some challenges such as proper IT support, lack of IT training for the midwives, lack of integrity between different software systems used in the midwifery centres and hospitals and attitude of the pregnant mothers. Based on our findings we provide some recommendations and design implications to support improved care provided by the midwives.

Keywords: Prenatal care, midwives, Information Technology (IT), maternal health, Human-Computer Interaction (HCI).

1 Introduction and Background

A key aspect to improve maternal health is to make use of the role of skilled health professionals, specifically midwives. Women around the world depend on the expertise of midwives for both prenatal and neonatal care. In the Netherlands, midwives are independent medical practitioners who work either in a midwifery practice centre or in a hospital. Approximately 78% of pregnant women start their maternity care with a primary care setting where 44% start labour in primary care, and finally 33% of women give birth under supervision of a primary care midwife [10]. Midwives are trained to provide maternity care to women whose pregnancy and childbirth are uncomplicated including antenatal, intrapartum and postnatal care to mother and child [1]. The first appointment at the midwifery practice usually takes place around the 8th week of pregnancy. On average one pregnant woman visits midwives 10-13 times over 9 months. Due to the frequent visits throughout pregnancy and because of their expertise, midwives are treated as dependable providers of pregnancy related health education and advice for pregnant women.

Maternity care in the Netherlands is known for its high percentage of home births and for the independent situation of midwives [13, 16]. In the Netherlands about 175,000 births occur annually [2]. Obstetric care in the Netherlands is provided by

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 635–642, 2013.

primary caregivers: midwives or general practitioners (GPs), and by specialist obstetricians. Women with low-risk pregnancies receive antenatal care from midwives. When they remain low risk throughout pregnancy they are free to choose to give birth either at home or in hospital. Their midwife (or GP) will support them in their choice and attend the birth in the chosen place. Referral to specialist care only happens in case of complications [1].

Each year worldwide 529000 maternal death occur [3]. The target set for MDG-5 is a 75% reduction of maternal mortality ratio between 1990 and 2015. Developing countries are mainly struggling to achieve the Millennium Development Goal (MDG). There are three main approaches of safe motherhood strategies which are developed based on different stages in a women's reproductive cycle. The stages are categorized as pregnancy, antenatal, delivery and postpartum period [7]. Antenatal care consists of a number of interventions administered to women during pregnancy, containing screening tests, immunizations, and treatment for recognized complications. It has been established that the detection of early signs or threat factors for morbidity and mortality can be spotted and effective interventions are possible [12]. The main aim of antenatal care is to produce a healthy mother and baby. Midwives have a significant impact on the wellbeing of women and new born babies.

Our aim is to identify needs and desires of midwives about the tools they need to deliver the best possible care for both mothers and infants. We need to get a clear picture of maternity care and their providers in developed countries though it is assumed that developed countries have better maternal care. There has been ample literature on midwives and maternal care in low resource environments for example see [7, 12, 15]. Furthermore, Human-Computer Interaction (HCI) community is working with midwives to design better systems such as improved ultrasound [6]. There are few studies which explored role of computer and Information Technology for midwives in developed countries. For example, one study conducted in Australia to understand the use of Internet among midwives [9]. It was revealed that midwives had insufficient Internet competence. We are not aware of any study that investigates workflow of midwives and the role of IT. Therefore, we explore maternity care issues and challenges in the Netherlands with respect to use of Information Technology (IT) in the midwifery centres. To better comprehend and support the workflows of midwives we conducted an interview study to discover the needs of midwives, and to develop technologies that will help them to better serve their patients. The specific objectives of the study are: a) identifying the role of IT and point-of-care testing (POC) devices b) identifying the purposes of Information Technology (IT) and POC for midwives, c) understanding the barriers to using IT and POC devices in the midwifery centres, d) recommendations to overcome these barriers to improve maternal care.

2 Interview Study and Results

In total 15 midwives participated in a semi-structured interview. They were contacted via the Royal Dutch Organization of Midwives (KNOV) [4]. The selection criteria were: the midwives should have active participation in midwifery centres and they should have minimum 5 years of working experience. The participation was voluntary. The preliminary questions for the interviews were a) key issues related to providing basic prenatal care, b) managing information to support prenatal care, c) prenatal visits and quality of care, d) barriers and facilitators to the prenatal care and role of Information Technology. Each interview session was one hour long. Interview data were audio recorded and later transcribed. The data were thematically analysed by the first author and one external researcher and the qualitative data analysis method described in [14] was followed. The key themes that emerged from the interviews are explained below:

Protocols for Midwives. Generally midwives (MW) follow written protocols in the midwifery centres; however midwives with several years of experience do not need those very often. One midwife (MW3) mentioned, "*some are written and some are in my head, it is in my head as I am experienced and for new colleagues it is on paper as well. We have several booklets in different languages*.'

Protocols are written by the Dutch organization of midwives (KNOV). It is a manual to inform how a usual prenatal care should be carried out. However, midwives are not forced to do that. One midwife (MW1) mentioned, '*You have your own responsibilities and you can decide based on your experiences and judgment when to see the patient in two week or four weeks. However, the rules are means and they are not aims'*. There are several protocols that are suitable for midwifery centres and regular hospitals where prenatal care is given as well. One midwife (MW9) mentioned, '*we work together with hospitals and we need to know the protocols of the hospital. We have not put the protocols on paper; they change a lot so it is in digital format as PDF files'*.

Patients and Communication. Midwives use both paper and electronic forms to manage patient information. During or before the first visit a prenatal registration form is given to the pregnant women and later those are used to keep online record with the help of the software systems that midwives use. One midwife (MW7) mentioned, "*when I take a blood sample for testing, I fill out a form, the results come back online and I decide if this is ok or not. If it is not OK the patients are referred. If I have to go an urgent consultation I call the obstetrician*". Furthermore, electronic agenda are quite often used to manage and provide care which helps them to work efficiently. Another midwife (MW5) mentioned, "*we have agenda per location and we also travel in the 5 locations to provide care so that pregnant women can easily reach. The agenda helps us to keep track of all the appointments"*.

During the first prenatal visit midwives collect the history of the patients and keep in their software systems. They also provide information to pregnant women about their practice and give lot of information about pregnancy such as diet, exercise and follow up visits Therefore, it is important that the provided information given to pregnant women are easy to remember and manage during the 9 months.

Managing Information and Supporting Prenatal Care. We identified that all the midwives complained about the software systems that they use to manage patient information. There are numerous major advantages of the software systems as well as reported by the midwives. The key advantages are a) patient summary information such as pregnancy history, disease history, family history are easily available, b) an agenda in the software system to manage appointments with the pregnant women.

The other things are, a) counselling information: smoking and drinking, b) provides summary of the health situation of the pregnant women based on self-reported data. One midwife (MW12) mentioned, *"If my colleague fills it I can just access the data. I always read the whole journal and it helps me to get an overall impression about a patient's condition"*. Therefore, the software systems help to maintain a good work-flow of midwives. Some key issues reported my midwives are, a) the software do not provide a notice if wrong information is added, b) the software systems do not let midwives know that the blood pressure is high. For example, one midwife (MW8) mentioned, "*the other program [name of the software program] gives me some jokes. If I type 11 children it gives me a joke. If the baby's blood type is O negative and we have to give medication. The software systems do not provide much intelligence. The whole thinking part is from the midwife. Though I believe that it cannot be fully taken over by the computer. The person who enters the data needs to take the decision"*.

Furthermore, the questionnaire data for prenatal registration collected from the pregnant women stays at the midwifery practice during the 9 months, when a baby is born and care has been finished, it goes to the national database anonymously [11]. For example, when someone has a miscarriage, midwife has to fill in a special form with all the information: there is no option in the software systems to do the same task. Midwives also use some point of care diagnostic devices such as blood pressure machine, Doppler machine, ultrasound in some practices and then sterilizing instruments are used for home delivery. These point-of-care (POC) testing devices are usually independent and each measurement data needs to enter into the software systems of the midwives.

Prenatal Visits and Quality of Care. The quality of care depends on the attitudes of the pregnant women and midwives. Sometimes some pregnant women have difficult behaviour. Some pregnant women do not even come for a follow up visit and they do not even keep their appointments up to date. One midwife mentioned (MW10), "*though we have a booklet for the pregnant women and we note down their follow up appointments, still they miss the appointments"*. In the regions where low social status people do not come to a prenatal visits regularly. Midwives need a lot of interaction with the pregnant women. Sometimes people do not show up and midwives have to call them. It would be nice to have a reminder system for those pregnant women. One midwife mentioned, "*the main part of our work is the relation with the pregnant women. If you do not have a good relation, you cannot give a good care. We do a lot of talking and listening. One of the basic competencies of midwives is making and building relations, maintaining relations and finishing the relations"*.

Barriers and Facilitators to Prenatal Care and IT. Four software systems are used in the midwifery centres such as a) MicroNatal (http://www.micronatal.nl/), b) Orfeus (http://www.orfeus.nl/), c) Onatal (http://www.onatal.nl/), and d) Vrumun (http://www.vrumun.nl/) and each midwifery centre uses only one of them. These software systems provide a wide range of support and generally the software system are secured and require secure log in credentials. One midwife (MW5) mentioned, "*we have a system everybody has different discipline, I do not know the finance section of the software so I cannot access that part. You get a user name and a special* *password to log on to the system"*. The positive aspect is that the software allows midwives to provide direct feedback to the developers and helps to provide better care to midwives by giving overview of the finished tasks. However, midwives had mixed reaction about the current software systems. One of the key issues is the broken link among the different software systems of midwives and gynaecologists. One midwife mentioned, "*I have my software and the gynaecologists have their own software. However, due to privacy reasons those are not connected. I want to check my patient within the system of the gynaecologists who is in a hospital"*.

Other wish list from the midwives were a) a touch program to input data while visiting a patient, b) easy customization of the menu items and other functional features where more frequent menus to be appeared only. One midwife mentioned, "*when I have to visit a patient who cannot come to our midwifery centre, I use a diary to keep notes and patient data. I cannot access the patient data which are in my computer at the midwifery centre. It is a pity and I have to enter the notes and other information when I am back at my office"*.

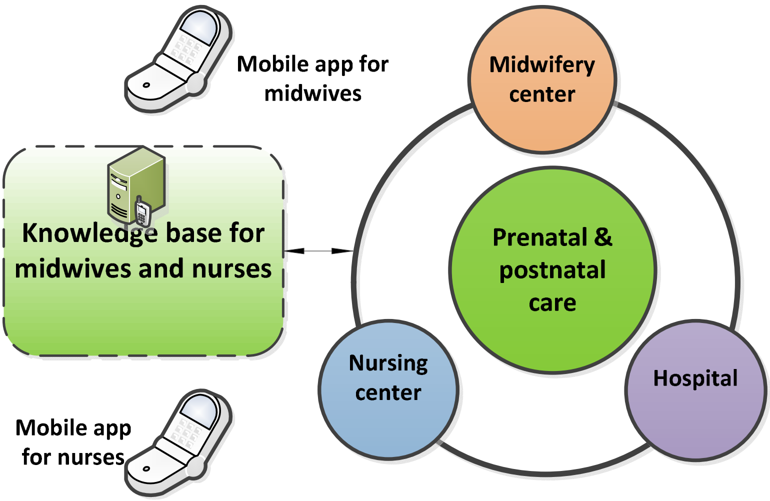


Fig. 1. Our proposed conceptual model to support improved prenatal and postnatal care. The keys issue is to repair the broken link among three care providers that is create an access to the patient data from each care location when required. Furthermore, mobile apps are proposed for midwives and nurses when they need to visit pregnant women and need to access the patient data. An integrated knowledge base is also proposed for improved care.

Midwives work closely with nurses (in Dutch they are called 'kraamzorg') and the nurses very often need critical information from midwives or specialists. For example, if a baby is born with high bilirubin which needs to be checked and very often nurses call midwives or specialists to know whether the bilirubin level is within normal range. One midwife mentioned (MW12), "*this is really critical information and an assistive mobile application would certainly help the nurses and midwives to take a decision and reduces their burden"*.

3 Recommendations and Design Implications

In this following section we provide recommendations to help midwives to provide better prenatal care.

Interoperability and Centralization of Software for Midwives. Too much information is provided to the pregnant women, therefore pregnant women either forget or they ask midwives repeatedly in upcoming visits. Therefore, it would be helpful if that information could be shared for both pregnant woman and midwives in an easy and accessible manner. For example, an interactive application/mobile app where it would be easy to interact with and pregnant women will get support over the nine months and post natal care would be really helpful.

It would be beneficial to have an integrated software system with doctors, midwives and other caregivers. The current system in the Netherlands does not allow different stakeholders to access the same patient data from their own software terminals. Therefore, creating an integrated health care system is a crucial need where privacy of the end user data will be preserved. The short term solution would be to make a viable link with the existing software systems operating in the midwifery centres and hospitals. We have observed that some of the software systems are not possible to use if the midwives need to visit a patient since the application can only be used at the computers of the midwifery centres. Therefore, a mobile application which could help to integrate with the local software system would be really helpful for the midwives. We recommend having a mobile app that would help to access the software systems while midwives have to visit a patient. The mobile app should be able to synchronise the gathered data later on.

Accessible Protocols. Since the maternal care system is based on protocols, and the obvious advantage would be to create some means to access the protocols based on the specific situation of pregnant women. Furthermore, creating a central knowledge base for the midwives would be helpful to reduce the burden for the new midwives.

Usability and Features of the Software for Midwives. We have found that there are four kinds of software systems and they have positive aspects while providing care by the midwives. All the software provides reasonable support to midwives despite some usability issues. The adoption process is based on price, and recommendations from other colleagues and the associated midwifery centres who work together. Improving the usability of the current midwifery software systems used in the midwifery centres will certainly help the workflow of the midwives. The major problems that we encountered were divided in two sides: a) lack of usable software system, and b) lack of financial support of the midwifery centres. One of the software was developed by individual initiative and no formal usability was measured. Furthermore, some manufacturers of the software systems started to get feedback from the midwives however since the community who use the systems are not very large the investment that is required to improve usability is less than the predicted return. We would like to stress that opinion of midwives should be taken into account while developing the software and support from the government is required for the software development firms.

Virtual Midwives. Two types of user groups were identified who ask care from the midwives. One group is motivated to visit the midwifery centre and the other group often misses their appointments. We also found that midwives find their task harder due to attitude of the people which is hard to control. However, it would be helpful if midwife could be virtually connected with the patients which could reduce their workload. Most of the time midwives are overloaded with so many queries of the women who are first time pregnant. Therefore, creating a virtual community and getting regular reminders about appointments could help to bridge the situation. It has also been observed that some user groups are very motivated and eager to know the growth of their expected baby. Since the midwives check the growth once in a month and sometimes it is hard to follow some appointments it would be beneficial to have a software system that will inform pregnant women about the growth of their baby. Furthermore, the will help pregnant women to get a piece of mind and will lower the burden from the midwives for example they will receive less worrying question from the pregnant women. It is evident from existing literature that the workload for primary care midwives are huge [17].

Training for Midwives. Providing regular training to midwives about the usage of IT or existing software systems will improve their workflow. We observed that some midwives are not well aware about the capability of the IT systems. Therefore, regular training to make them competent will result in delivering improved care. These findings are also in line with the existing findings mentioned in [5, 8].

Tools to Support Postnatal Care. One of the issues midwives mentioned that they have problems in postnatal care support as well. The main reason was that the nurses involved in postnatal care are not sure about different measurements such as high bilirubin of the new born child. Most of the time the nurses call the midwives or doctors to consult. In some cases they have to calculate the bilirubin in every two days and calling nurses or doctors is very cumbersome. Therefore, a mobile application where nurses could measure the level of severity of the bilirubin would help to improve their workflow.

4 Conclusion and Future Work

In this paper we have presented a qualitative study to understand the work process and tools used to provide maternity care by midwives. The findings are encouraging though the prenatal care in the Netherlands is considered very good and some of the findings are due to the nature of the Dutch healthcare system. We are working with local midwives to design improved services for midwives and pregnant mothers. For example, we are designing an app for the peace of mind of pregnant mothers. The objective is to help pregnant mothers to get useful information and make them aware about the growth of the baby over 9 months. The app will also be able to document each visits to the midwifery centres and will provide recommendations on food and nutrition, exercise etc. Moreover, we believe that in order to bridge the different care providers, privacy of the information was an issue. KNOV is working to solve the issues so that each stakeholder can access each other's data to provide better pre and post natal care. In this study, we have interviewed the midwives from 12 different midwifery practices in the Netherlands. The experiences of the midwives varied in terms of software use, prior familiarity with IT and training. However, the findings led us to discover a better information system to improve maternal care in the Netherlands which may be suitable in similar other countries as well.

Acknowledgments. We thank the midwives and KNOV for their cooperation.

References

1. Anthony, S., Amelink-Verburg, M.P., Jacobusse, G.W., van der Pal-de Bruin, K.M.: De thuisbevalling in Nederland 1995-2002 (Home birth in the Netherlands 1995-2002) Bilthoven/Leiden, Stichting Perinatale Registratie Nederland, TNO Kwaliteit van Leven (2005)
2. Bais, J.M.J., Pel, M.: The basis of the Dutch obstetric system: risk selection. European Clinics in Obstetrics and Gynaecology 2(4), 209–212 (2006)
3. Bullough, C., Meda, N., et al.: Review: Current strategies for the reduction of maternal mortality. BJOG: An International Journal of Obstetrics & Gynaecology 112(9), 1180–1188 (2005)
4. Dutch midwifery association, `http://www.knov.nl`
5. Gerrish, K., Morgan, L., et al.: Factors influencing use of information technology by nurses and midwives. Practice Development in Health Care 5(2), 92–101 (2006)
6. Havukumpu, J., Vähäkangas, P., Grönroos, E., Häkkinen, J.: Midwives experiences of using HMD in ultrasound scan. In: Proceedings of the 4th Nordic Conference on Human-Computer Interaction: Changing Roles, NordiCHI 2006, pp. 369–372. ACM (2006)
7. Herz, B.K., Measham, A.R.: The safe motherhood initiative: Proposals for action, World Bank (1987)
8. Hillan, E.M., McGuire, M.M., et al.: Computers in midwifery practice: a view from the labour ward. Journal of Advanced Nursing 27(1), 24–29 (2001)
9. McKenna, L., McLelland, G.: Midwives' use of the Internet: an Australian study. Midwifery 27(1), 74–79 (2011)
10. Manniën, J., Klomp, T., et al.: Evaluation of primary care midwifery in the Netherlands: design and rationale of a dynamic cohort study (DELIVER). BMC Health Services Research 12(1), 69 (2012)
11. Netherlands Perinatal Registry (PRN foundation), `http://www.perinatreg.nl`
12. Rooney, C.: Antenatal care and maternal health: how effective is it? A review of the evidence Maternal Health and safe Motherhood Programme division of Family Heath. WHO, Geneva (1992)
13. Smeenk, A.D.J., ten Have, H.A.M.J.: Medicalization and obstetric care: an analysis of developments in Dutch midwifery. Medicine, Health Care and Philosophy 6(2), 153–165 (2003)
14. Strauss, A., Corbin, J.M.: Basics of qualitative research: grounded theory procedures and techniques. Sage Publications (1990)
15. The state of world's midwifery, `http://www.unfpa.org/sowmy/resources/en/main.htm`
16. Wiegers, T.A., van der Zee, J., et al.: Variation in home birth rates between midwifery practices in the Netherlands. Midwifery 16(2), 96–104 (2000)
17. Wiegers T.A: Workload of primary-care midwives. Midwifery 23, 425–432 (2007); 7. NVOG: Nota

Technology Enhanced PBL in HCI Education: A Case Study

Christina Vasiliou, Andri Ioannou, and Panayiotis Zaphiris

Cyprus University of Technology
Limassol, Cyprus
{c.vasiliou,andri.i.ioannou,panayiotis.zaphiris}@cut.ac.cy

Abstract. Problem Based Learning (PBL) is an instructional method in which the base for learning is a real-world problem. A typical PBL setting is comprised of students working together on an authentic problem, using simple tools such as whiteboards and stationery. Online tools and multimedia technologies have also been used to support PBL activities. There is however no empirical work on the blended use of both physical and digital tools. This paper presents a case study in which we employed PBL pedagogy for the teaching of a post-graduate course in Human Computer Interaction (HCI). The activities were situated in a multimodal information space, rich in digital and physical elements including personal computers, projectors with downwards projection, tablets, iPods, digital pen readers, stationery and a Facebook group for each team. We administrated questionnaires assessing students' motivational beliefs and overall satisfaction with the learning experience. Our results show that students' overall satisfaction was highly rated, while the information space contributed to students' engagement and collaboration.

Keywords: Problem based learning, HCI education, Multi-modal interfaces.

1 Introduction

A common barrier in acquiring knowledge presented in class is that students are often disengaged from active collaboration, interaction and reflection. Problem based learning (PBL) is a didactic approach in which the base of learning is the need to solve a real-world problem [1] and providing context-rich situations [2]. A typical PBL setting is comprised by students working together on the problem, using simple stationery. With the advancement of the technology new ideas were brought forefront. Online tools such as forums and blogs, as well as multimedia enriched settings, including games, have been used to support PBL. There is however no empirical support on using both physical and digital tools together in a PBL setting. Specifically, this study attempts to achieve a two-fold purpose by integrating PBL in a technology enhanced environment: first, to explore its possibilities and constraints for teaching HCI; and second, to explore how this space promotes learners' collaboration and engagement.

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 643–650, 2013.

2 Literature Review

As this research project focuses on the use of PBL pedagogy and the use of technology, this section will present a review of traditional PBL as well as technology enhanced PBL case studies.

2.1 Problem Based Learning

PBL is an instructional method in which, problem is the trigger that initiates learning; while students work in small collaborative groups to solve it [3]. Therefore, students take the responsibility of their own learning objectives and new information is expected to come from their investigation towards the solution of the problem. A traditional PBL setting is comprised by students working together on the problem using simple physical elements such as markers, whiteboard and post-it notes.

2.2 Use of Technology in PBL

Online Tools and PBL Environments. Initial forms of technology included Internet purely for online research or online tools and environments to facilitate PBL. Several forms of synchronous and asynchronous communication tools and online environments have been identified among studies, such as WebCT [4], Blackboard [5], Blogs and Wikis [6]. Results showed that students preferred to communicate real time due to the immediate type of communication simulating physical interaction [6].

With these results, a blended model of PBL was implemented integrating face-to face and online interactions [7]. Through the literature analysis, we identified two approaches, based on the leading party. Studies where the leading role is retained by the online environment embraced the need for physical meetings, requiring a number of physical meetings either for progress reflection and feedback [8] or for face-to-face PBL sessions [9][10]. In the second approach, face-to-face PBL sessions are followed by the use of online tools for communicating, sharing and reflecting [11][12][13]. Another study [14], enhanced digital elements in class with the use of an interactive whiteboard to support group collaboration. Overall, with blended environments, students felt more willing to get involved in learning and felt satisfied with the flexibility in time for online tools [13].

PBL within Multimedia, Games and Virtual Worlds. The second main pathway in the use of technology in PBL is multimedia. Researchers created multimedia enriched environments successfully increasing students' motivation [15][16]. Due to their popularity [17], games were also embraced by PBL community, with efforts made to understand game elements that can lead to the improvement of educational technologies, enhancing students' engagement [18]. Redesigned courses were highly rated by students, but no significant changes were found in terms of learning.

What appears to be a tendency lately is the use of 3D virtual worlds for the development of virtual spaces where learners could engage in problem solving, waving the confrontment of their decisions' consequences [19]. One of the most widely used 3D environments in educational settings is Second Life (SL), mainly due to its abilities of building rich 3-D environments, highly immersive for players with a relatively low cost to participate [20]. In brief, studies showed that 3-D worlds offered a fun and engaging approach towards learning, while the involvement of professionals and the context of future tasks in a realistic setting increased their motivation [21].

Identifying a Gap in Knowledge. Online and multimedia environments have all been extensively used in PBL studies, but little attention has been paid to the blend of digital and physical elements during in-class activities. Results were positive towards engaging students through the blended form of in-class activities [14], raising the following questions: Does the interaction between a variety of physical and digital elements in the PBL environment promote learners' engagement and collaboration? To address this, our study set out to develop a multimodal information space, including both digital and physical elements and to further investigate how this space promotes students' engagement and collaboration.

3 Methods

This research aims at developing a blended multimodal information space and evaluating the levels of engagement and collaboration. The following section demonstrates the case study, describing in detail the blend of physical and digital elements, the developed information space and the data collection process.

3.1 Participants

The sample was composed of 30 postgraduate students (11 males and 19 females) enrolled in an HCI course. Students' background was multidisciplinary including studies such as computer science and games, graphic arts and multimedia or education and communication. The participants were divided into five groups (5 groups of 5 to 7 students), each formed with students from different disciplines.

3.2 Procedures

PBL Curriculum and Course Structure. A traditional lecture based Human Computer Interaction (HCI) post-graduate course was reformed to fit PBL. The HCI course was organized in 3 hour weekly sessions, including mini lectures and group activities.

The mini lectures were intended not to directly provide content, but form students stimulus to progress with their own research and acquire knowledge on their own, matching the PBL pedagogy[2]. As Hmelo-Silver indicated "a lecture at the right time may be beneficial", denoting that in the process of solving a problem it might be necessary to provide certain information to the learner to support self-directed learning.

Group activities aimed to enable student engagement and active collaboration within each group were usually two hours long and consisted of three main units:

(1) Weekly Reflection: A 20 minutes session to summarize and reflect on what has been done since the last laboratory session. (2) Brainstorming: Usually 80 minutes session to discuss about the progress of the project, readjust the problem based on the new facts and discuss ideas. (3) Self Study Assignment: A 20 minutes session to define what are the groups learning objectives and form a to-do list.

Problem. The problem given to students for the design project was derived from CHI2013 student design competition scenario, entitled “Empowering the Crowd: Changing Perspectives Through Collaboration”[22]. Problems are of crucial importance to the success of PBL [1]. The selection of the current problem scenario was based on the need for a challenging, real world problem.

Setting. The blended delivery format of PBL enclosed a combination of traditional face to face and online interactions, rich in digital and physical tools. Its main features include whiteboard, post-it notes, markers and other stationery, typical for traditional PBL and expanding it with a set of digital tools. More specifically, a ceiling-mounted projector for downwards projection, one or more personal computers, one tablet, one iPod, one sense camera and an infrared pen and pen reader (IPPR) per group. The structure of the multimodal information space used in our study is shown in Figure 1.

Small technologies and gadgets, such as iPods, tablets and stationery, were grouped and organized in a box, constructing our “Creativity Box” (CB). A detailed guide was developed and distributed to the students during the first week laboratory with the purpose of learning how to use each element. Students were instructed to use the provided technologies freely, with no limitations on how and when to use each element. Figure 1, demonstrates individual use of digital elements included in the information space. Facebook was used as the medium for synchronous and asynchronous interactions allowing participants to communicate and share material. Due to scope and space limitations of this paper, Facebook data analysis and results will be presented elsewhere.

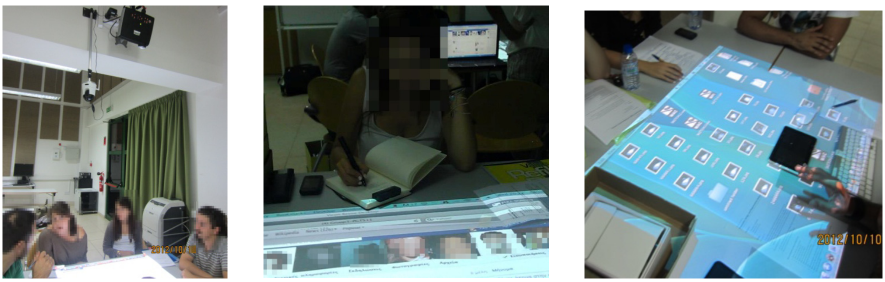


Fig. 1. Left: Structure of the multimodal information space. Center: Taking digital notes. Right: Downward projection as a shared screen.

3.3 Data Collection

Questionnaires were constructed, mainly with the aim of gauging students’ views of how the information space setup supported their collaborative work. All questionnaires were administrated after the full duration of the course and included Likert-type items with a 7-point agreement response scale (from 1: completely disagree to 7: completely agree).

The first questionnaire concerned students' motivational beliefs, to measure three constructs of interest: (1) Motivation, measuring task value and academic self efficacy, (2) Technology Expectations and (3) Technology Self-Efficacy. The second questionnaire was dealing with students' collaboration experience based on five constructs of interest: (1) Communication and Interaction, (2) Reflection, (3) Perceived Learning, (4) Overall Satisfaction, and (5) Frustration.

Open ended questions were given to the participants concerning the way each group functioned throughout laboratory activities, such as "Did you assign roles?", "Did you split the work?" and addressing issues on teamwork and the setting.

4 Results

Upon the completion of the research study, all the recorded data were analyzed with mixed methods. The current section presents an overview of the results from the post questionnaires and open ended questions.

4.1 Post Questionnaires

Due to missing data, results only from 29 students were reported and analysed. First the internal consistency for each subscale was assessed using Cronbach's alpha; all 9 subscales had acceptable internal consistency (Cronbach's alphas > .70). Afterwards, all subscale mean scores were calculated. As demonstrated in Table 1, subscale mean scores were all above the midpoint of the response scale suggesting that participants found their activities useful and important (M=6.02, SD=.75) and felt confident with the use of technology (M=5.89, SD=.74). As demonstrated in Table 2, mean scores were above the midpoint of the response scale for the first four measures, and way below midpoint for Frustration, suggesting that educational technology and collaborative activities were positively embraced by the students. The participants indicated that overall their collaborative learning experience satisfaction was high (M=6.01, SD=.87) and that the activity was successful in promoting the construction of knowledge (M=6.25, SD=.75), critical thinking and reflection (M=5.79, SD=.93).It is also important to denote that even though frustration levels were clearly below the midpoint of the response scale, the variations in answers among the students was high (M=2, SD=1.22).

Table 1. Subscales and Descriptive Statistics for Post Questionnaire on Motivation (N=29)

Subscale	# Items	Cronbach's Alpha	Means (SD)
1. a) Task Value	6	.890	6.02 (.75)
b) Academic Self-Efficacy	6	.893	5.64 (.65)
2. Technology Expectations	6	.963	5.57 (1.07)
3. Technology Self-Efficacy	6	.926	5.89 (.74)

Table 2. Subscales and Descriptive Statistics for Post Questionnaire on Collaboration (N=29)

Subscale	# Items	Cronbach's Alpha	Means (SD)
1. Communication and Interaction	6	.855	6.16 (.62)
2. Reflection	5	.908	5.79 (.93)
3. Perceived Learning	5	.776	6.25 (.75)
4. Overall Satisfaction	5	.927	6.01 (.87)
5. Frustration	4	.789	2.00 (1.22)

4.2 Open-Ended Questions

Qualitative feedback was in line with previous findings in that students were positive overall concerning the functioning of their group and the course structure. In terms of course structure and collaborative work, students stated that "It was a wonderful experience!", or "Great and effective way of organizing the lesson". In terms of the information space participants supported the use of technologies in class supporting that "Technologies provided in class helped us". The majority of the participants further reported good collaboration and communication between team members, equality in group work and satisfaction with the final product, suggesting high levels of agreement and collaboration within the group.

On the negative side, a few participants found the use of some technologies was frustrating or not really useful. This finding is in line with questionnaire results that indicated low levels of frustration but with great variations, suggesting that few participants found the developed information space more challenging than others or some certain technologies not as useful as others. This is something we seek to investigate further in future work.

5 Discussion

This study set out with the aim to develop a multimodal information space, including both digital and physical elements and to further investigate how this space enhances learners' collaboration and engagement. The results of the questionnaires and open ended questions indicate that the developed information space was positively perceived by students, engaging and motivating them towards learning.

Regarding our first goal to investigate the blended form of the multimodal information space, our findings indicate that the blend was successful but improvement areas have been identified. Similarly to [14], students embraced the blended mode of instruction with increased satisfaction as demonstrated from questionnaires. But, high variations in frustration levels, based on questionnaires and different opinions from qualitative feedback, show us that elements included in the information space, especially digital tools, should be further investigated.

On the question of whether the developed information space enhanced students' collaboration and engagement the study found that to be true. Based on the questionnaire results, participants demonstrated high levels of communication and interaction and found class activities useful and valuable. Furthermore, our results lend support to previous studies indicating that engagement and motivation are highly rated in technology enhance PBL environments or curriculums, especially when including multimedia [16]. Students' exhibit enthusiasm and positive attitude towards the course and group activities as identified within the qualitative feedback. All these are indicators of students' engagement and collaboration in the process of learning, answering the second part of our research aim.

Although the research has reached its two-fold target, a number of important limitations need to be considered. This research was conducted on a small size population attending the HCI course; therefore, results cannot be generalized. Another potential limitation of this study is not measuring learning through this multimodal information space to successfully identify it as a learning space.

6 Conclusion

This paper presented a case study in which we employed PBL pedagogy for the teaching of a post-graduate course in HCI. Our results show that students' overall satisfaction was highly rated, while the information space contributed to students' engagement and collaboration. The evidence from this study supports that digital and physical elements can be well blended for the support of the HCI, PBL classroom.

Acknowledgement. This project is funded by the Cyprus Research Promotion Foundation (DESMI 2009-2010) and the Slovenian Research Agency (ARRS), under the "Bilateral Cooperation" between Slovenia and Cyprus (ΔΙΑΚΡΑΤΙΚΕΣ/ΚΥ-ΣΛΟ/0411).

References

1. Hung, W., Jonassen, D.H., Liu, R.: Problem-based learning. In: Spector, J.M., van Merriënboer, J.G., Merrill, M.D., Driscoll, M. (eds.) Handbook of Research on Educational Communications and Technology, 3rd edn., pp. 485–506. Erlbaum, Mahwah (2008)
2. Hmelo-Silver, C.E.: Problem-based learning: What and how do students learn? Educational Psychology Review 16, 235–266 (2004)
3. Barrows, H.S.: Problem-based learning in medicine and beyond: A brief overview. New Directions for Teaching and Learning, 3–12 (1996)
4. Schell, R., Kaufman, D.: Critical Thinking in a Collaborative Online Multimedia PBL Simulation. In: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education, pp. 448–455. AACE, Chesapeake (2007)
5. Gürsul, F., Keser, H.: The effects of online and face to face problem based learning environments in mathematics education on students' academic achievement. Procedia - Social and Behavioral Sciences 1, 2817–2824 (2009)
6. Lo, H.-C.: Utilizing Computer-mediated Communication Tools for Problem-based Learning 12, 205–213 (2009)
7. Voos, R.: Blended learning – what is it and where might it take us? Sloan-C View 2, 2–5 (2003)
8. Woltering, V., Herrler, A., Spitzer, K., Spreckelsen, C.: Blended learning positively affects students' satisfaction and the role of the tutor in the problem-based learning process: results of a mixed-method evaluation. Advances in Health Sciences Education: Theory and Practice 14(5), 725–738 (2009)

9. Hoic-bozic, N., Mornar, V., Boticki, I., Member, S.: A Blended Learning Approach to Course Design and Implementation 52(1), 19–30 (2009)
10. King, S., Greidanus, E., Carbonaro, M., Drummond, J., Boechler, P., Kahlke, R.: Synchronous Problem based learning e-Learning (e-PBL) in Interprofessional Health Science Education. Journal of Interactive Online Learning 9(2), 133–150 (2010)
11. Donnelly, R.: The nexus of problem-based learning and learning technology: Does it enable transformative practice? European Journal of Open, Distance and E-Learning (2009)
12. Donnelly, R.: Harmonizing technology with interaction in blended problem-based learning. Computers & Education 54(2), 350–359 (2010)
13. Lou, S.J., Shih, R.C., Tseng, K.H., Diez, C.R., Tsai, H.Y.: How to promote knowledge transfer through a problem based learning Internet platform for vocational high school students. European Journal of Engineering 35(5), 539–551 (2010)
14. Bridges, S.M., Botelho, M.G., Tsang, P.C.S.: PBL 2.0: Blended Learning for an interactive problem-based pedagogy. Medical Education 44(11), 1131 (2010)
15. Liu, M., Toprac, P., Yuen, T.T.: What Factors Make a Multimedia Learning Environment Engaging: A Case Study. In: Zheng, R. (ed.) Cognitive Effects of Multimedia Learning, pp. 173–192. Information Science Reference, Hershey (2009)
16. Liu, M., Olmanson, J., Horton, L., Toprac, P.: Motivational Multimedia: Examining Students' Learning and Motivation as They Use a Multimedia Enriched Learning Environment (2011)
17. Walker, A., Shelton, B.E.: Problem-based educational games: Connections, prescriptions and assessment. Journal of Interactive Learning Research 19(4), 663–684 (2008)
18. Warren, S.J., Dondlinger, M.J., McLeod, J., Bigenho, C.: Opening The Door: An evaluation of the efficacy of a problem-based learning game. Computers & Education 58(1), 397–412 (2012)
19. Parson, V., Bignell, S.: Using Problem-Based Learning Within 3D virtual worlds. Cutting-edge Technologies in Higher Education 4, 241–261 (2011)
20. Warburton, S.: Second Life in higher education: Assessing the potential for and the barriers to deploy virtual worlds in learning and teaching. British Journal of Educational Technology 40(3), 414–426 (2009)
21. Omale, N., Hung, W.C., Luetkehans, L., Cooke-Plagwitz, J.: Learning in 3-D multiuser vir-tual environments: Exploring the use of unique 3-D attributes for online problem-based learning. British Journal of Educational Technology (2009)
22. CHI student Design Competition, `http://chi2013.acm.org/authors/call-for-participation/student-competitions/student-design-competition/`

The 'Relay Ideation' Technique: Moving from Problem Understanding to Problem Solving in the Design Process

Karin Slegers[1], Pieter Duysburgh[2], and An Jacobs[2]

[1] CUO | Social Spaces, KU Leuven / iMinds, Parkstraat 45, bus 3605, 3000 Leuven, Belgium
[2] iMinds – Digital Society, SMIT – VUB, Pleinlaan 9, 1050 Brussels, Belgium
karin.slegers@soc.kuleuven.be,
{Pieter.Duysburgh,An.Jacobs}@vub.ac.be

Abstract. When describing the design process in product innovation, many authors identify phases that can be described as 'problem analysis' and 'generating ideas'. Several techniques are available to support design teams in each of these phases, but it remains a challenge to move from understanding a problem to coming up with ideas for concepts that might solve the problem. In addition, some of these techniques have counterproductive social side effects, which in fact may impede creativity in a design team. In this paper we describe a new technique for product idea generation called the 'relay ideation' technique. This technique was developed to help design teams move from understanding a problem to thinking creatively and concretely about the problem in order to generate concepts for innovative products or services. The technique is illustrated with a case study about IT applications for hearing-impaired children.

Keywords: Ideation, techniques, conceptualization, design team.

1 Introduction

Various authors have tried to capture and understand the creative process in design and have identified several steps in this process (for an overview see [1]). The majority of authors have identified two phases that can be understood as 'problem understanding' and 'problem solving'. In this respect, much has been written on involving targeted end-users and design teams in general in the design process (e.g. [2]), and particularly on how design concepts emerge from their sources of information (e.g. [3]). In addition, an abundant number of techniques has been developed and documented to support design teams in these phases of the design process (see [4] and [5] for some examples), many of which are variations on Osborn's 'brainstorming' technique [6]. In order to solve problems, this technique focuses on quantity, refrains from criticism, encourages unusual ideas and combines and improves ideas.

The transition from understanding a problem to solving it, which can be particularly challenging, can be understood as moving from openness to closure [7]. First, a design team needs to be open for ideas and views of others. Subsequently, it needs to draw conclusions and formulate solutions for observed needs and problems. Of course, in design thinking, 'a problem' and 'a solution' mean something different than

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 651–658, 2013.

in formal logic [8]. While in formal thinking a solution is either right or wrong, in design thinking this cannot be said: no answers can be considered 'correct'. In design thinking, solutions are to be understood as 'proposals', 'ideas', and 'suggestions'. Just as a 'the problem' is situational, and refers to the designer's current understanding of the situation, a 'solution' in design is temporal and always an ongoing process [8].

An additional challenge in the process of ideation and finding solutions, is the fact that design teams often struggle with group effects that are counterproductive for idea generation [1]. Theoretically, problem solving with a team should result in more creative solutions compared to an individual approach. But research has extensively shown that in practice the opposite is true. For instance, Warr and O'Neill [1] have identified three social influences on creativity that impede the theoretical benefits of group creativity: (a) *production blocking*, or the fact that verbally expressed ideas in a group allow only one person to express ideas at one time, while subsequently other team members may forget or suppress their ideas; (b) *evaluation apprehension*, or the fear of criticism of other team members; (c) *free riding*, or the phenomenon of group members becoming lazy and relying on other team members to contribute.

In this paper we present the relay ideation technique. With this technique we tried to find a means to help a design team (a) move from understanding to solving a problem and (b) overcome social group influences in the team that may impede creativity. Central to this technique is the concept of empathy. As the quality of design solutions is related to the empathy of a design team with the target group, the likelihood that products and services fit the end-users' lives well increases when these users and their experiences are well understood [9]. Therefore, before generating new ideas, a design team needs to gain empathy with the targeted user group, in order to understand the problem and to come up with relevant and useful solutions.

In what follows we explain the relay ideation technique. We further illustrate this technique by means of a practical research case. Finally, in the discussion, we will hint at some aspects of the technique we wish to explore further.

2 Analysis and Ideation

The relay ideation technique aims to facilitate the transition between the two phases in the creative process we mentioned above: problem analysis and problem solving.

2.1 Problem Analysis

It is assumed that before using the relay ideation technique, there has been a preparatory phase in which insights have been gathered about the needs and wants of targeted end-users, as is the common first step in a user-centered design process [10]. For the relay ideation technique, increasing empathy is seen as an essential step in analyzing the problem. In user-centered design, often only a few members of the design team (often design or social researchers) are directly involved in studying the needs and wants of the targeted end-users. However, when thinking about solutions for the users' problems, a larger group of people with various backgrounds is often involved (e.g. computer science, design, engineering, domain expertise). All members of the team need to gain empathy with the targeted users to find real solutions for them. In our view, simply communicating user insights to the design team is insufficient to let the design team think creatively about the needs and wants of the end users. Therefore, we recommend doing a team empathy exercise.

One example of such an empathy exercise is to have the design team create a number of what we refer to as ‘empathy characters’. Although such empathy characters resemble personas that are often created in the design process to represent the problems and needs of the target group [11], they are not based on extensive ethnographic research and do not necessarily truthfully represent all characteristics and needs of the target group. Instead, empathy characters are a means to help members of the design team to empathize with the target group. In the empathy exercise, members of the design team bring a very basic character description to live by using user insights they were presented by the design or social researchers (we describe this process further in the case study in section 3). By doing this, the design team is encouraged to actively process the insights, which helps them to increase their understanding of the users. The team members discuss their empathy characters with each other and relate them to the project the team is working on. As such, several problems and challenges tend to come up that are relevant for the target group and for the project they are working on. A final step in the problem analysis phase is to select a number of these problems for the problem solving phase.

2.2 Problem Solving: The Relay Ideation Technique

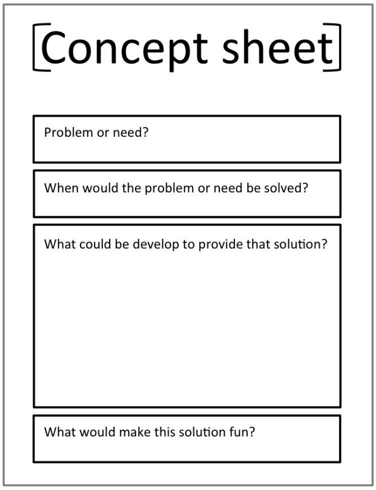


Fig. 1. The concept sheet

It is the problem solving phase that requires the design team to move from the problems and challenges analyzed previously to ideas for solutions. For this purpose, the relay ideation technique (illustrated in Fig. 1) encompasses an approach of working together in couples to ‘solve’ the problems selected in the problem analysis phase. This is facilitated by using so-called ‘concept sheets’. These concept sheets consist of four empty fields that need to be completed in order to think of problem solutions in a stepwise fashion. The couples are not asked to fill in all the fields at once, but to complete only one field, and then pass the concept sheet on to the next couple. Ideally, there are at least as many couples as there are empty fields on the sheet.

In the first step of this problem solving procedure, each couple selects one of the problems that were listed in the problem analysis phase and describes more details of this problem in the first field of the concept sheet. After passing on the concept sheet to the next couple, each couple thinks of a general, abstract solution to the problem they were ‘given’ by another group. After passing on the concept sheets again, each couple formulates characteristics and functionalities of a concept that would provide this solution. Finally, after passing on the concept sheets one last time, the couples think of how to make the concept attractive.

Once all the fields of the concept sheets are completed, the exercise has resulted in a number of concepts equal to the amount of couples participating to the session (since every couple was given a concept sheet at the start). Each couple presents the concept they have finalized to the whole design team. The concepts can be further discussed, and others might add additional ideas or insights, and make adjustments to the concept. If time and attention allows, the exercise can be repeated.

3 Case Study: Ideation for Hearing-Impaired Children

We have used the relay ideation technique to generate new concepts for IT research projects in several contexts now, including assistive technology, music education and public transportation. To further illustrate the use of the relay ideation technique, we will briefly elaborate upon one of the projects in which we feel the technique was successfully used [12]. This project aimed at developing innovative IT concepts for hearing-impaired children, but besides that, the project had no specific technological focus at the start and was still in a very open, unfocused stage at the time the relay ideation technique was used. It was a collaborative project involving several academic partners. In the case study, two creative workshops were organized with the project design team, consisting of social researchers, designers and engineers. The workshops were moderated by the two social scientists who had also done the preparatory ethnographic research. The other team members had specific technical or design skills, but no knowledge of the targeted users.

3.1 Problem Analysis: Empathy Exercise

In the first half-day workshop, the two researchers who had done the ethnographic study communicated the main insights they had gained at a small poster ‘exhibition’. The team was then subdivided into five couples and each couple was given an empathy character template. This template contained some basic demographic facts of the character the couples had to create, and four assignments that stimulated the couples to bring their character to life: (a) describe a typical day of this person (b) think of the most important problems and needs that this person might experience; (c) list this person’s likes and dislikes in life; (d) describe possible dreams this person might have. For the basic demographics, four types of hearing-impaired children were included in the templates to reflect the diversity of the targeted users. Two characters were completely deaf; two were hard of hearing. One of the deaf children had a hearing family; the other had a deaf family. Two of the four children went to a special school; one of them lived at the school during the week.

By creating these empathy characters, all couples had to actively process the information they were presented with in the poster exhibition and they had to apply the information to the life of the particular character they were given. While filling in the template, the couples could ask additional questions to the moderators or consult the posters. When finished, the couples presented their characters to the entire design team. The moderators made use of this moment to correct some misunderstandings of the design team with respect to the insights that were first presented. Meanwhile, the problems and challenges that the characters were described to encounter in their lives were shared with all team members and listed on a flipchart.

The moderators felt that his empathy exercise was quite successful and helped the design team to better understand the targeted end-users. A problem that could be noted, however, was that a certain bias might result from repeatedly reinterpreting the original data. In all cases, the original data had been interpreted twice: once by the field researchers, and a second time by the design team. When a sign language interpreter had been involved in the ethnographic phase, the information from the hearing-impaired children had even been interpreted a third time. Even though the field researchers corrected some misunderstandings at the end of the character exercise, it is quite likely that their understanding of the target group was biased too.

3.2 Problem Solving: Relay Ideation

The aim of the second workshop, also a half-day workshop, with the design team was to generate ideas for new IT applications that might solve some of the problems that hearing-impaired children face with. This workshop took place in the same room as the first workshop. The posters of the exhibition, the empathy characters and the list of problems and needs were still in place and as such still visible to the team. The team was split up into four couples, this time for a relay ideation session.

During this relay ideation session, the concept sheets were used as described above. Each couple started by choosing and detailing one of the problems or needs listed on the flipchart in the first workshop. Then the concept sheets were passed on and the couples tried to describe a (non-technical) solution for the problem that was chosen by the previous couple. Again, the sheets were passed on and each couple tried to think of a concrete concept for a new IT application that might provide the solution described on the template they just received. Finally, after another exchange of concept sheets, each couple reflected on the solution and the concepts by indicating what elements would provide a fun experience for the children. As such, in one hour, 13 concepts were generated (one sheet included two concepts).

This process of circulating a step-by-step concept sheet turned out to be a thought-provoking technique. Because each couple had to think of different elements of the concepts (problem, solution, IT concept, fun elements), participants mentioned afterwards that the task never became boring. They felt continuously spurred and stimulated, because each couple constantly received surprising input from another couple. Neither was the exercise too difficult, as the couples did not have to think the whole solution through: they could just think of the next step and then literally handed the rest of the problem over to the following couple.

After the exercise with the concept sheets, the moderators added an additional task in which each couple tried to design a creative artifact made of scrap material, representing one of the concepts. To this aim, each couple chose one of the concepts that appealed most to them. By encouraging them to use creative materials, we wanted to stimulate the participants to 'think-by-doing', which affords unexpected realizations that one might not have arrived at without creating a concrete artifact (or prototype) [13]. As such, the couples became more critical of the concept and could reflect on it in more detail in terms of functionality, interaction, design, etc.

The concepts that resulted from the relay ideation approach in this project were validated in a workshop with technical project members and in two workshops with teachers of hearing-impaired children. Finally, three concepts were deemed to hold enough potential for further research and development.

4 Discussion

The relay ideation technique subdivides the transition from understanding a problem to solving a problem in small, incremental steps. By distributing the responsibility for solving the entire problem by passing on the concept sheet after finishing one single step of the problem solution process, we hoped to facilitate creative thinking. As such, the team members don't have to think the whole problem through at once. We wanted the relay ideation technique to be a productive technique: depending on the time reserved for each field on the concept sheet to be completed and the number of groups participating, the exercise should quickly result in a considerable amount of ideas. Finally, co-creating a concept may result in a feeling of shared ownership and support for the concept by the whole design team, which may in return increase the team involvement and the likelihood of further research and development on the concept.

This feeling of shared ownership might to a certain extent compensate for evaluation apprehension, one of the impeding social factors that has been linked to group brainstorms [1] that we mentioned in the introduction. As a result, criticism is less likely to be concentrated on a single person. Also, the fact that each field was completed by a couple, makes it less likely that a single person is to 'blame' for a step on the concept sheet. As the relay ideation technique allows for asynchronous working, it may also compensate for production blocking, another impeding social factor. Every couple is continuously busy, thinking about the field on the concept sheet they have in front of them. Furthermore, the fact that the team is split up into couples who have to pass on their work to the next team repeatedly, minimizes free riding, or the risk of any participant being lazy or relying on other team members too much. Working in small teams of two persons and (similar to the 'group passing technique' or the '6-3-5' method [5]) rotating the concept sheet creates a balance between the two opposing dynamics of restricting identification to avoid criticism and enabling identification to avoid the free riding effect suggested by Warr and O'Neill [1].

As with any idea-generating technique, the relay ideation technique does not guarantee high quality ideas. The quality depends on the participants, their inspiration, their motivation and many other determinants that are beyond the control of a technique. Therefore, the concepts that result from the relay ideation exercise should be regarded as ‘raw materials’, ideas that should be reviewed critically and will only to some extend be useful for the project.

The role of criticism in the ideation process has been a topic of much discussion. Osborn [6] believed criticism impedes creativity, and that it should be kept separate from thinking of new ideas. Warr and O’Neill [1] have also referred to fear of criticism as an impeding factor. However, research has shown that criticism is actually also essential to move beyond predictability and enhance creativity in idea generation [14]. For that reason, we included a moment to present and discuss the concepts that were generated at the end of the workshop. This can be done in a discussion, but also by, for instance, using scrap material, to make the concept more concrete and thus closer to reality (see e.g. [13] for more on engaging tangible working materials).

The relay ideation technique, in combination with the preparatory empathy exercise, is an attempt to stimulate group creativity by minimizing the effect of several contradictory social psychological effects. As we have not yet been able to validate the technique, in our future research we plan to experimentally compare the effectiveness of the techniques with that of other idea generation methods. In addition to this, we plan to study the experience of the members of the design team using the relay ideation technique into more detail.

Since its development, we have used the technique at several occasions. In future research projects we plan to develop and validate the technique further by experimentally testing which elements of the technique objectively result in better ideation processes. The experiences we have had with the technique so far already gave us several ideas to further alter and improve the technique:

(1) While an empathy exercise is an essential part of the relay ideation technique, there is not always time to allow for an elaborate process of context mapping and familiarizing the team with the end-users and their context. Further iterations of our technique should explore how this process can be more efficient, e.g., by including the character template in the concept sheet.

(2) Many models do not see creativity as a linear process [15]. However, our technique forces the design team to solve a problem in a linear manner. For new version of the relay ideation technique we’d like to explore possibilities to facilitate a non-linear approach to problem solving. Exactly how this may be achieved is still unclear.

(3) Relay ideation can be used in several stages of the design process. In the case presented here, we used the technique in what is often described as the ‘fuzzy front end’ of design [2]. In later phases, the design team needs to take more issues and project decision into account. These restraints can influence the ideation process, resulting in different types of ‘solutions’, or requiring the design team to deal with the generated ideas in another manner.

In future research we hope to further develop the technique, which is still rather experimental and in that sense a work-in-progress, along these lines.

Acknowledgement. This work was part of the iMinds GR@SP project funded by iMinds (Interdisciplinary institute for Technology), a research institute founded by the Flemish Government.

References

1. Warr, A., O'Neill, E.: Understanding design as a social creative process. In: Proceedings of the 5th Conference on Creativity & Cognition, pp. 118–127. ACM, New York (2005)
2. Sanders, E.B.-N., Stappers, P.J.: Co-creation and the new landscapes of design. CoDesign 4, 5–18 (2008)
3. Halskov, K., Dalsgaard, P.: The emergence of ideas: the interplay between sources of inspiration and emerging design concepts. CoDesign 3, 185–211 (2007)
4. Herring, S.R., Jones, B.R., Bailey, B.P.: Idea Generation Techniques among Creative Professionals. In: 42nd Hawaii International Conference on System Sciences, HICSS 2009, pp. 1–10 (2009)
5. Bonneux, A., De Couvreur, L., Michiels, P., Vanneste, C., Verthé, B.: Innowiz, `http://www.innowiz.be/index.php`
6. Osborn, A.F.: Applied imagination: principles and procedures of creative problem-solving. Creative Education Foundation, Buffalo (1993)
7. Steen, M.: The fragility of human-centred design (2008), `http://www.marcsteen.nl/texts.htm`
8. Löwgren, J., Stolterman, E.: Thoughtful Interaction Design: A Design Perspective On Information Technology. MIT Press (2004)
9. Koskinen, I., Battarbee, K., Mattelmäki, T.: Empathic Design: User Experience in Product Design. IT Press (2003)
10. ISO/IEC: 9241-210:2010 - Ergonomics of human-system interaction - Part 210: Human-centred design for interactive systems, `http://bit.ly/YEhJ05`
11. Rosson, M.B., Carroll, J.M.: Usability engineering: scenario-based development of human-computer interaction. Morgan-Kaufmann, New York (2002)
12. Slegers, K., Duysburgh, P.: Research Methods for Involving Hearing Impaired Children in IT innovation. In: Proceedings of NordiCHI, Reykjavik, Iceland, October 16-20, pp. 781–784 (2010)
13. Agger Eriksen, M. Engaging design materials, formats and framings in specific, situated co-designing: A micro-material perspective. In: Online Proceedings of the Third Nordic Design Research Conference, Oslo, Norway (2009)
14. Nemeth, C.J., Personnaz, B., Personnaz, M., Goncalo, J.A.: The liberating role of conflict in group creativity: A study in two countries. European Journal of Social Psychology 34, 365–374 (2004)
15. Howard, T.J., Culley, S.J., Dekoninck, E.: Describing the creative design process by the integration of engineering design and cognitive psychology literature. Design Studies 29, 160–180 (2008)

The Effect of Stress on Cognitive Load Measurement

Dan Conway[1,2,*], Ian Dick[1,3,*], Zhidong Li[1,4], Yang Wang[1], and Fang Chen[1,4]

[1] NICTA, NSW 2015, Australia
[2] University of Newcastle, NSW 2308, Australia
[3] University of Sydney, NSW 2006, Australia
[4] University of New South Wales, NSW 2052, Australia
ian.dick01@gmail.com, jujuspacejazz@hotmail.com,
{zhidong.li,yang.wang,fang.chen}@nicta.com.au

Abstract. Human physiological signals have been widely used to non-invasively measure cognitive load (CL) during task execution. A major challenge for CL detection is the presence of stress, which may affect physiological measurements in ways that confound reliable detection of CL. In this experiment we investigated the effect of stress on cognitive load measurement using galvanic skin response (GSR) as a physiological index of CL. The experiment utilized feelings of lack of control, task failure and social-evaluation to induce stress. Mean GSR values were shown to be significantly different between CL levels in the 'no-stress' condition, but not when including the 'stress' condition. On the other hand, features extracted from GSR signals based on peak detection exhibited consistent behaviour under both conditions, demonstrating the usefulness of the features as cognitive load index even when a person's stress level is fluctuating.

Keywords: Cognitive load, galvanic skin response (GSR), stress.

1 Introduction

The term cognitive load (CL) refers to the amount of cognitive resources required for a person to complete a certain task. CL has been shown to have important implications for learning [13], safety in driving [4], aviation [16], and user interface design [10]. Being able to accurately monitor CL in real world environment would have important and far-reaching implications. This capability could enable the real-time detection of cognitive overload, when a person's cognitive resources are overloaded with information, and this could help to avoid dangerous situations, such as an air traffic controller with too many concurrent responsibilities. CL monitoring could also enable the design of better computer interfaces, since interface designs could be rated based on the CL levels they impose on users, and designs with poor performance could be disregarded.

* First authors.

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 659–666, 2013.

Physiological signals have previously been proposed as a method of quantifying CL. Some notable successes in CL evaluation have been achieved via signals such as respiration, brain wave, and pupillary response [6] [14] [15] [17] [18]. Another physiological signal that has been used successfully to quantify CL, and which is the focus of this research effort, is galvanic skin response (GSR), a measure of the electrical conductance of the skin shown to be sensitive to both CL and stress [5] [9].

Any given physiological signal used as an index for CL is likely to be affected by various additional inputs from the human body. Thus, a major task of cognitive workload measurement via physiological means is demonstrating the diagnosticity and construct-validity of the CL index. One of the major contributors to change in human physiological systems is stress [8]. Stress has been shown to effect both the sympathetic and parasympathetic nervous systems and, in its more extreme states, results in large changes to physiological function that may well obscure the relationship between a physiological indicator and CL. Furthermore, stress may, in some circumstances, be a confounding factor for CL in that changes in CL may correlate with changes in stress levels. Construct-validity must be established before we can safely assert that changes in physiological indicators are the result of CL and not stress or other confounding factors. This experiment studies the effect of stress on CL measurement using GSR.

Although definitions of stress vary, there is good consensus in the literature regarding conditions where it is likely to arise [3] [7] [11]. Failure at a task, together with feelings of lack of control, in situations where participants are evaluated by others is a widely used paradigm for stress induction. These conditions could be operationalised in an experimental paradigm developed by Dedovic et al. called the ‘Montreal Imaging Stress Task’ (MIST) [2]. The experiment presented here closely follows the MIST protocol with minor operational adjustments.

2 Related Work

GSR has been used successfully in the past to index CL. In an experiment involving traffic control management it was illustrated that the mean GSR of test subjects increases as the difficulty of cognitive tasks increases [12]. In addition, [9] analysed the time and frequency domains of recorded GSR signals and showed that CL imposed through arithmetic and reading tasks can be indexed by GSR of test subjects.

The relationship between GSR and stress has also been examined. In an experiment involving driving tasks, [5] were able to successfully classify different driving periods based on the stress levels of the driver. They extracted useful features from GSR signals recorded during the experiment based on peak detection and input these features into machine learning classification algorithms with positive results. [19] combined GSR with several other physiological signals to classify the stress states of test participants through the use of machine learning tools. Stress levels were induced by having subjects complete a “Paced Stroop Test”, where the colour of a word that spells a different colour must be identified.

An interesting study on discriminating stress from cognitive load was carried out by Setz et al. [11]. However, they did not experimentally manipulate CL and only demonstrate the ability to differentiate between 'stress' and 'no-stress' conditions, where CL level was consistent between the two conditions. Nonetheless, the feature detection processes they outlined appear promising, and have been further studied in the experiment presented here.

3 Experiment

11 male students and employees (24-49 years' old, ten right handed and one left-handed) took part in the experiment. Participants were offered one movie ticket and biscuits as recompense for their participation.

All participants had the voluntary nature of the experiment explained to them and then filled out a paper version of the Kessler K-10 Psychological Distress Scale [1] to ascertain that they were unlikely to be vulnerable to ongoing negative effects from the stress condition. Only participants who scored less than 19 (thus fell into the category 'likely to be well') were permitted to continue the experiment. Three potential candidates were rejected via this means.

3.1 Apparatus

All experimental stimuli were presented on a computer screen using custom software whilst participants were sitting comfortably at a desk. GSR signals were collected using a Thought Technology 'ProComp Infiniti' interface and its 'SCFlex/Pro' skin conductance sensor. The sensors were attached to the non-dominant hand for all participants. GSR signals were sampled at a rate of 10Hz. The brain wave signals were also recorded with a low-cost EEG device 'Emotiv EPOC'. However the analysis of the EEG data is out of the scope of this paper. Participants were asked to remain still and only move their dominant hand for mouse control during the experiment.

3.2 Procedure

The experiment consisted of a within-subjects, six-way factorial design. There were math questions of three difficulty levels (low, medium and high) administered under two different stress conditions: 'no-stress' and 'stress'. For level 1 problems (low difficulty), three terms were added together. For level 2 (medium difficulty), each problem consisted of four terms, with both addition and subtraction required. Level 3 problems (high difficulty) consisted of five terms, with addition, subtraction and multiplication required. The multiplication terms were in a random position within the problem.

All participants undertook the 'no-stress' condition first. Participants were told that they would be completing math tasks but it was emphasized that their performance/accuracy was not important. After submitting some basic demographic information, a two minute baseline period was carried out where the participants were told,

Fig. 1. Experiment setup

via an on screen prompt, that they should just relax and let their mind wander. Then three two-minute blocks of math tasks were presented, each with 4 multiple choice answers. Tasks were not time limited and feedback was not provided. The three blocks in the ‘no-stress’ condition were of level 1, 2 and 3 difficulty in sequential order. The participant was given a two-minute resting period in between each block.

After block three, the participants were asked, via on screen prompts, to nominate a ‘target score’ for further tasks based on their estimation of their performance so far. Once submitted, the stress condition ensued. They were told that their performance would be now be monitored. They were also informed of time limits for further trials, and video screens were switched on so that the test subject could see a video of themselves and also of other people observing them (see Figure 1). Now in the stress condition, three more blocks of level 1, 2 and 3 math tasks were carried out, again with two-minute pauses in between each block but with time limits now imposed on each trial. Feedback (‘correct’, ‘wrong’ or ‘out of time’) was provided for one second after each trial.

4 Analysis

In order to test the validity of the methods used in the experiment for inducing different levels of cognitive load, one-way ANOVA test of pooled subjective ratings was conducted. Results showed that the difference between the different difficulty levels was significant ($F = 82.32$, $p < 0.05$). It can be seen in Figure 2 that the means of each group were increasing with the task difficulty level.

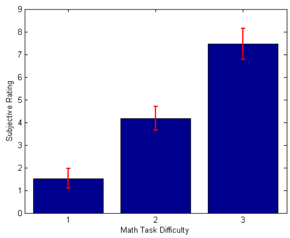
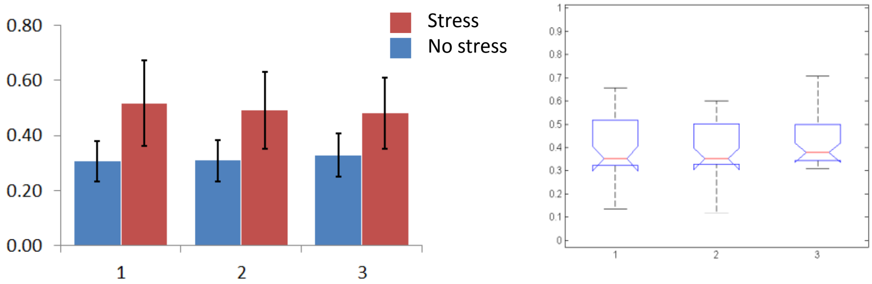


Fig. 2. Subjective rating of task difficulty

4.1 Analysis Using Mean GSR

The mean GSR values were inspected to study the effect of stress on cognitive load measurement. The distribution of normalised mean GSR values corresponding to the sub-sections of math task difficulty 1, 2 and 3 under both 'no-stress' and 'stress' conditions can be seen in Figure 3a. To investigate the relationship between mean GSR and cognitive load when no stressful stimuli are present, we conducted ANOVA analysis on these GSR values under 'no-stress' conditions and found that there are statistically significant differences between the 3 different levels ($F = 10.5$, $p < 0.05$), and there is a noticeable upward trend in mean GSR that corresponds to an increase in task difficulty.



(a) (b)

Fig. 3. Distribution of normalized mean GSR values for task difficulty levels 1,2,3 under the 'no-stress' and 'stress' conditions

The positive correlation between cognitive load and GSR could no longer be observed, however, once the data from the 'stress' part of the experiment are included for consideration. Figure 3b shows the distribution of the normalised mean GSR values for math task difficulty levels 1, 2 and 3, with both the 'no-stress' and 'stress' data included. ANOVA analysis of these values does not produce significant results ($F = 0.05$, $p = 0.95$). The results indicate that mean GSR cannot effectively index cognitive load when stress levels are fluctuating, since mean GSR is sensitive to stress and the correlation between cognitive load and mean GSR becomes obfuscated when stress is a confounding factor. To overcome this problem, feature extraction technique has been employed for workload evaluation in the following section.

4.2 Analysis Using 'Peak' Features

Similar to [5], several features corresponding to the peaks in the signals were extracted from the smoothed GSR signals. The following definitions were made: S_D is the distance along the x-axis from the local min preceding a peak to the local max of the peak (i.e. peak duration); S_M is the distance along the y-axis from the local min preceding a peak to the local max of the peak (i.e. peak magnitude); S_F is the number of peaks divided by the task period (i.e. peak frequency). Figure 4 illustrates these

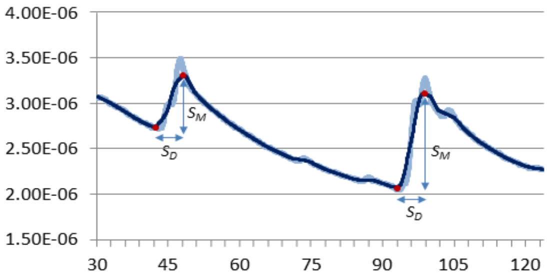
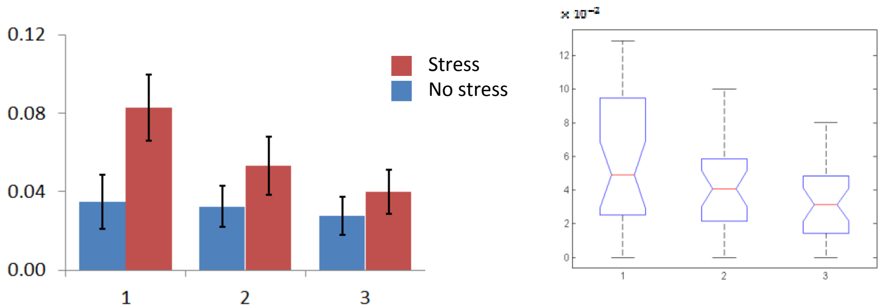


Fig. 4. Example of a smoothed GSR signal adorned with S_D and S_M features

concepts. In this work we used these three peak based GSR features to study the effect of stress on cognitive load measure. Except S_M, the other two features, S_F and S_D, demonstrated their usefulness for indexing CL even when stress is a confounding factor.

The S_F feature represents the frequency of peaks per sub-section. The 'no-stress' and 'stress' distributions for this feature are shown below in Figures 5a. For mean GSR, there was no common trend between the 'no-stress' and 'stress' conditions, and so there is no obvious way of using mean GSR to index CL when stress levels are fluctuating. In contrast to this, the S_F feature displayed in Figure 5 behaves similarly in both the 'no-stress' and 'stress' conditions. We can see that S_F is negatively correlated to task difficulty regardless of whether stress is present, although the relationship is stronger in the 'stress' condition. Figure 5b shows the distribution of the S_F feature for difficulty levels 1, 2 and 3, with both the 'no-stress' and 'stress' data included. ANOVA analysis was performed on this data to test the significance of the negative correlation. The result exhibited significant difference among the three difficulty levels ($F = 3.96$, $p < 0.05$).

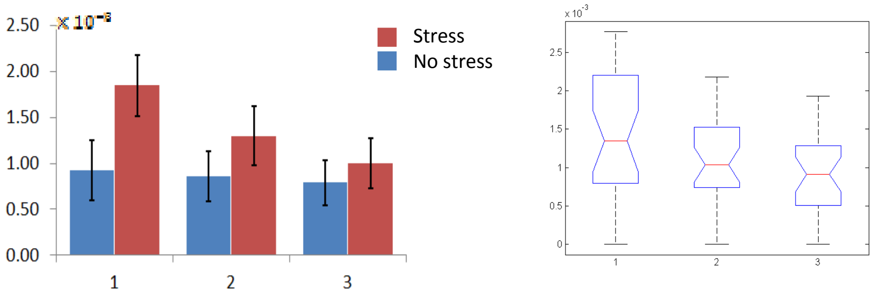
The S_D feature corresponds to the peak duration per sub-section. The distribution of normalized S_D feature (sum of peak durations divided by the sub-section period)



(a) (b)

Fig. 5. Distribution of S_F feature for task difficulty levels 1, 2, 3 under the 'no-stress' and 'stress' conditions

corresponding to math task difficulty 1, 2 and 3 under both ‘no-stress’ and ‘stress’ conditions can be seen in Figure 6a. It turns out that this feature behaves quite similarly to the S_F feature, and is negatively correlated to task difficulty under both ‘no-stress’ and ‘stress’ condition. Figure 6b shows the distribution of the S_D feature for difficulty levels 1, 2 and 3, with both the ‘no-stress’ and ‘stress’ data included. The downward trend with increasing task difficulty could be observed. ANOVA analysis also generate significant difference among the three difficulty levels ($F = 5.14$, $p < 0.05$), indicating the feature could be useful as an index of cognitive load even under the influence of stress conditions.



(a) (b)

Fig. 6. Distribution of normalized S_D feature for task difficulty levels 1, 2, 3 under the 'no-stress' and ‘stress’ conditions

5 Conclusion

The experiment in this work helped to reinforce GSR as an index of cognitive load during task execution. Without the impact of stress, it appears that an increase in CL (induced by increasing the difficulty of tasks given to test subjects) results in an increase in mean GSR value. This relationship is, however, obfuscated when test subjects experience fluctuating levels of stress. Stress was introduced into the experiment using an adaptation of the MIST protocol, and this blurred the connection between GSR and CL.

GSR may still be useful as an index for CL even when stress is a confounding factor, if we consider peak based features extracted from the GSR signal other than the mean value. Both peak frequency in the signal and peak durations are negatively correlated to task difficulty and hence CL. These features could possibly be used to dissociate CL from stress and develop a stress-agnostic method of CL classification. Our future work in this area include more feature extraction and also machine learning to investigate whether these features may be used to classify a person’s CL under the influence of various confounding factors.

References

1. Andrews, G., Slade, T.: Interpreting scores on the Kessler psychological distress scale (K10). Australian and New Zealand Journal of Public Health 25, 494–497 (2007)
2. Dedovic, K., Renwick, R., Mahani, N.K., Engert, C., Lupien, S.J., Pruessner, J.: The Montreal Imaging Stress Task: using functional imaging to investigate the effects of perceiving and processing psychosocial stress in the human brain. Journal of Psychiatry & Neuroscience 30(5), 319–325 (2005)
3. Dickerson, S.S., Kemeny, M.E.: Acute Stressors and Cortisol Responses: a theoretical integration and synthesis of laboratory research. Psychological Bulletin 130(3), 355–391 (2004)
4. Engstrom, J., Johansson, E., Ostlund, J.: Effects of visual and cognitive load in real and simulated motorway driving. Transportation Research Part F: Traffic Psychology and Behaviour 8(2), 97–120 (2005)
5. Healey, J., Picard, R.: SmartCar: detecting driver stress. In: International Conference on Pattern Recognition, vol. 4, pp. 218–221 (2000)
6. Knoll, A., Wang, Y., Chen, F., Xu, J., Ruiz, N., Epps, J., Zarjam, P.: Measuring cognitive workload with low-cost electroencephalograph. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) INTERACT 2011, Part IV. LNCS, vol. 6949, pp. 568–571. Springer, Heidelberg (2011)
7. Markus, H.: The effect of mere presence on social facilitation: an unobtrusive test. Journal of Experimental Social Psychology 14(4), 389–397 (1978)
8. Martin, G.N., Carlson, N.R., Biskit, W.: Psychology. Pearson Education, Harlow (2007)
9. Nourbakhsh, N., Wang, Y., Chen, F., Calvo, R.A.: Using galvanic skin response for cognitive load measurement in arithmetic and reading tasks. In: Proceedings of the 24th Australian Computer-Human Interaction Conference, pp. 420–423. ACM (2012)
10. Saadé, R.G., Otrakji, C.A.: First impressions last a lifetime: effect of interface type on disorientation and cognitive load. Computers in Human Behaviour 23(1), 525–535 (2007)
11. Setz, C., Arnrich, B., Schumm, J., La Marca, R., Tröster, G., Ehlert, U.: Discriminating stress from cognitive load using a wearable EDA device. IEEE Transactions on Information Technology in Biomedicine 14(2), 410–417 (2010)
12. Shi, Yu., Ruiz, N., Taib, R., Choi, E., Chen, F.: Galvanic skin response (GSR) as an index of cognitive load. CHI Extended Abstracts on Human Factors in Computing Systems (2007)
13. Sweller, J.: Cognitive load theory, learning difficulty, and instructional design. Learning and Instruction 4(4), 295–312 (1994)
14. Veltman, J.A., Gaillard, A.W.K.: Physiological workload reactions to increasing levels of task difficulty. Ergonomics 41(5), 656–669 (1998)
15. Wang, W., Li, Z., Wang, Y., Chen, F.: Indexing cognitive workload based on pupillary response under luminance and emotional changes. In: Int'l Conf. Intelligent User Interfaces (IUI), pp. 247–256 (2013)
16. Wilson, G.F.: An analysis of mental workload in pilots during flight using multiple psychophysiological measures. The International Journal of Aviation 12, 3–18 (2002)
17. Xu, J., Wang, Y., Chen, F., Choi, H., Li, G., Chen, S., Hussain, S.: Pupillary response based cognitive workload index under luminance and emotional changes. CHI Extended Abstracts on Human Factors in Computing Systems (2011)
18. Xu, J., Wang, Y., Chen, F., Choi, E.: Pupillary response based cognitive workload measurement under luminance changes. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) INTERACT 2011, Part II. LNCS, vol. 6947, pp. 178–185. Springer, Heidelberg (2011)
19. Zhai, J., Barreto, A.: Stress detection in computer users through non-invasive monitoring of physiological signals. Blood 5 (2008)

The PEW Framework for Worth Mapping

Fatoumata Camara[1], Gaëlle Calvary[2], and Rachel Demumieux[3]

[1] Objet Direct – Viseo, 4 avenue Doyen Louis Weil – 38000 Grenoble, France
[2] Grenoble INP, LIG, CNRS, BP 53 – 38041 Grenoble cedex 9, France
[3] Orange Communications SA, Rue du Caudray 4 – 1020 Renens 1, Switzerland
fcamara@objetdirect.com, Gaelle.Calvary@imag.fr,
Rachel.Demumieux@orange.ch

Abstract. In Human Computer Interaction, it is more and more clear that usability is not enough. In order to take into account the other criteria that may be relevant for design, G. Cockton introduced the notion of "worth" and the Worth Centered Design (WCD) framework for its operationalization. The WCD framework structures the development process and provides designers with a set of tools, including Worth Maps (WMs).

Worth maps connect systems attributes to human ones, and as such represent a promising tool. However, they remain understudied and under-experimented.

This paper presents the results of our experience with WMs. More precisely, it proposes the PEW (Perceived and Expected Worth) framework for worth mapping, reports findings from a study conducted with 5 experts regarding many aspects of WMs, and discusses future directions for research.

Keywords: Interactive systems design, worth, Worth Maps (WMs).

1 Introduction

The need to go beyond traditional criteria (e.g., usefulness, usability, learnability) in interactive systems design has been identified since years. For instance, the notion of emotional usability was formulated by Logan already in 1994 [3].

With the advent of ambient computing, this concern is becoming more and more important. Indeed, nowadays, the user is provided with the opportunity to interact with a large number of systems serving different purposes, using different kinds of devices, in different settings. This diversity of types of systems and of contexts of use has then given good opportunities to point other criteria relevant to consider in interactive systems design. For instance, from interviews that followed the field trial of the "Whereabouts clock" (a physical device that displays family members current location) elements such as reassurance, connectedness, expression of identity, and social touch emerged [9].

More recently, many works have investigated additional criteria for interactive systems design through new notions such as the new usability [4], the User eXperience

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 667–674, 2013.

(UX), or worth. The work presented here is related to the last aforementioned notion, worth, initially introduced as ‘value’ by G. Cockton in 2004 [6], as well as the WCD framework for worth operationalization. The WCD framework structures the development process around 4 phases [8]: (1) study of needs, wants, and unfelt needs (for worth elicitation), (2) design (for the system design and implementation), and (3) evaluation (for worth accomplishment) that my lead to iteration. G. Cockton also proposes a set of tools, including Worth Maps, for supporting the WCD.

From a theoretical point of view, the WCD appeared to be well suited for the design of next generations of interactive systems. For instance, in [5], the authors argue that “the framework is particularly relevant for designing pleasurable, enjoyable and entertaining interaction.” However, despite having this potential, the WCD has not received much attention from community. Indeed, the relevant literature regarding worth and the WCD framework has been essentially produced by Cockton (and colleagues). This notice motivated our work.

Our final goal is to promote the WCD through lessons of experience from a complete operationalization of the framework and the development/improvement of tools for supporting the method. In an earlier work, we have investigated the worth of a system named Cocoon (study of needs, wants, and unfelt needs). This work allows us a better understanding of the notion of worth. In a further step, we have engaged in worth mapping sessions (prior to design and implementation) for the so-called system in collaboration with five experts involved in interactive systems design. This paper reports our experience with worth maps. Outcomes from this experience are twofold: (1) we propose a framework for worth mapping, so called the PEW (Perceived and Expected Worth) framework; (2) we provide insights regarding WMs representation, construction, and interest.

2 Background Literature

2.1 Worth

Worth is anything that motivates the user to buy, learn, use, or recommend an interactive product, and ideally most or all of these [8].

The WCD begins by the identification of worthies and supports several starting points (e.g. field study [11], sentence completion [12]). In order to study the worth of Cocoon, a mobile and context-aware system that automatically pushes personal information (e.g. related to a maternity where a relative was born) and impersonal information (e.g. an historical building) to the user, we conducted 19 semi-directed individual interviews using storyboards. In order to surface elements beyond traditional criteria considered in Human-Computer Interaction (HCI) design, interviews focus not only on the “what” (features) and slightly the “how” (interface and interaction) but also on the “why” (motivations). Outcomes allowed a further understanding of worth.

- Worth is multidimensional

Outcomes from interviews revealed that worth is related to several aspects. Most of classes of elements identified from the analysis of existing WMs (see section 2.2) emerged from interviews: features (e.g. contextual information push), qualities (e.g. easy to use), consequences from usage (e.g. discovery of new places and new stories), and worthwhile outcomes (e.g. social impacts: maintain of ties). Interviews also revealed that associations of the aforementioned elements may activate the user' personal values such as family, friendship, freedom, and so on.

- Worth is twofold

During interviews, participants mentioned:

- The positive side of Cocoon (i.e. what would motivate them to buy, learn, use, or recommend the system). We called this positive side: the "*perceived worth*".
- The negative side of Cocoon (i.e. what would not make them buy, learn, use, or recommend the system): aversions, missing or poorly addressed features. We called **elements that compensate** this negative side: the "*expected worth*". For example, the two following verbatim are examples of aversions that the possibility of controlling the number of notifications from Cocoon compensates: "*I don't want to receive information everywhere otherwise I would have too much information*"; "*Can you imagine? I would be notified every time I pass near a building in order to inform me that a relative worked here*."

Intuitively, there is no expected worth in a perfect system and the goal of design should be to transform the expected worth to perceived worth.

- Worth is related to the system

As we mentioned above, worth is also related to personal values. However, interestingly, interviews revealed that it is tied to the system from a user's point of view. Consequently, "worth" is "worth of the system."

2.2 Worth Maps

WMs are inspired from Hierarchical Value Maps (HVMs) [2] used in Marketing to study customer motivations for purchase. An HVM is a diagrammatic integration of separately elicited Means-End-Chains (MEC) [1]. A MEC links a product attributes (concrete/abstract) to its desirable and undesirable consumption consequences (functional/psychosocial) that are in turn linked to values (instrumental/terminal).

WMs revisit HVMs in several points which are detailed in [11]. In HCI, WMs connect system-oriented attributes to user-oriented ones, thus shifting from "*designing as crafting to designing as connecting*" [12]. From an analysis of existing WMs, we have identified the following classes of system-oriented attributes: materials, features, qualities (of features); and the following classes of user-oriented attributes: usage consequences (i.e. actions, feelings, UX), worthwhile outcomes, aversions, and adverse outcomes.

WMs follow a vertical representation. The system-oriented attributes connected to user-oriented ones are placed above aversions and adverse outcomes. However, it is important to note the negative side (aversions and adverse outcomes) are not generally shown on WMs because they are supposed to be overcome by the upper side elements: "*the benefits are worth the costs*". Thus, a WM models an ideal system.

WMs support design in different ways. First, they represent a means to ensure that the system satisfies users' expectations (*credibility* [13]). Second, WMs maintain the design team focused on the design goals (*commitment* [13]) relevant evaluation criteria can be instantiated according to associations highlighted in WM [10]. As such, WMs represent a support for evaluation.

3 Worth Mapping for Cocoon

3.1 Individual Interviews with Experts

An initial version of the framework was produced on the basis of results from the study of worth of Cocoon and an analysis of existing literature regarding HVMs and WMs. This framework and an instantiated WM for Cocoon were then submitted to five people involved in interactive systems design as support for discussions during semi-structured individual interviews.

Interviews were conducted with: a project manager (7 years of experience), a User Interface (UI) and interaction designer (9 years of experience), a psychologist (12 years of experience), a graphic designer (15 years of experience), and a software engineer (11 years of experience). Participants worked for the research and development department of a telecommunication company and are used to work together within agile development teams.

Each interview lasted approximately 45 minutes. Interviews followed two goals: (1) assessing the understandability and the interest of WMs, (2) assessing the understandability and applicability of our framework. In order to reach these goals, interviews were structured around the following themes:

- role(s) of the participant in the design process;
- communication means and supports used in the participant's design team;
- systems evaluations goals;
- understandability and interests of WM (using Cocoon WM as example for illustration purpose);
- understandability and applicability of the worth mapping framework (using the initial framework).

3.2 Results

Interviews were recorded using a Dictaphone and analyzed using a speech analysis. Results are summarized below.

- <u>The framework for mapping Perceived and Expected Worth</u>

The resulting framework, so called the PEW framework, (see Fig.1) is only slightly different from the initial one. Indeed, regarding this aspect, participants' suggestions were mostly related to the form rather than to the content.

The major contribution of our framework is in reflecting the actual system (instead of an ideal one), at a given time (from a user's point of view) by making explicit both positive (perceived worth) and negative (expected worth) associations. This introduces implicitly another use of WMs as support to evaluation based on comparison of WMs representing the system at different design stages.

Our framework also suggests a vertical representation for WMs. Following existing representation patterns, the framework suggests representing positive associations (perceived worth) up and negative ones (expected worth) down. The Native Software and Hardware Components (NSHC) of the device hosting interaction may support features: they represent a key class of elements for WMs. As they are shared by both perceived and expected worth, they naturally come in the middle. It is important to note the linkage among the user-oriented attributes is simplified on the proposed framework: a hierarchy may exist (see Fig.2).

Participants of interviews investigating WMs use and interest for interactive systems design suggest that boxes borders and links width corresponds to their importance. Thus, relevant elements can be visually identified at first glance.

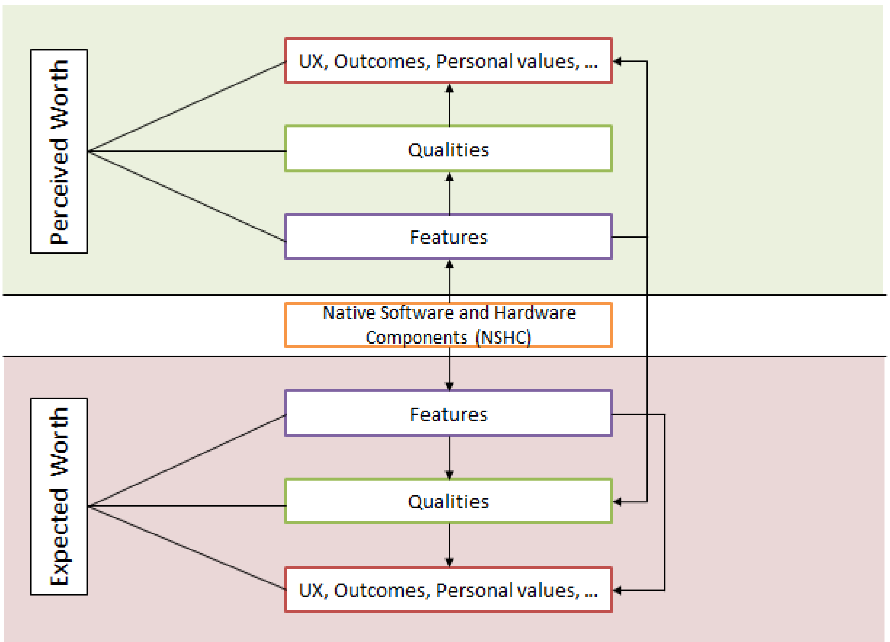


Fig. 1. The PEW framework

- WMs construction

Regarding construction of WMs two points appeared to be critical: worth elicitation and the base support for worth mapping. For instance, the psychologist asked whether WMs should be constructed manually from scratch (on the basis of worth elements). According to her, the worth mapping process may require too much time to be applicable in realistic design processes. Indeed, the study of users' deep motivations requires a careful analysis. Moreover, our experience shows that the construction of WMs from scratch is a fastidious task.

Participants then proposed the ontology for worth associating already elicited worth elements to different types of system, device, user, or features as well as tools that support construction of WMs. They also suggested that WMs are interactive and the developer suggested involving users in the construction of WMs.

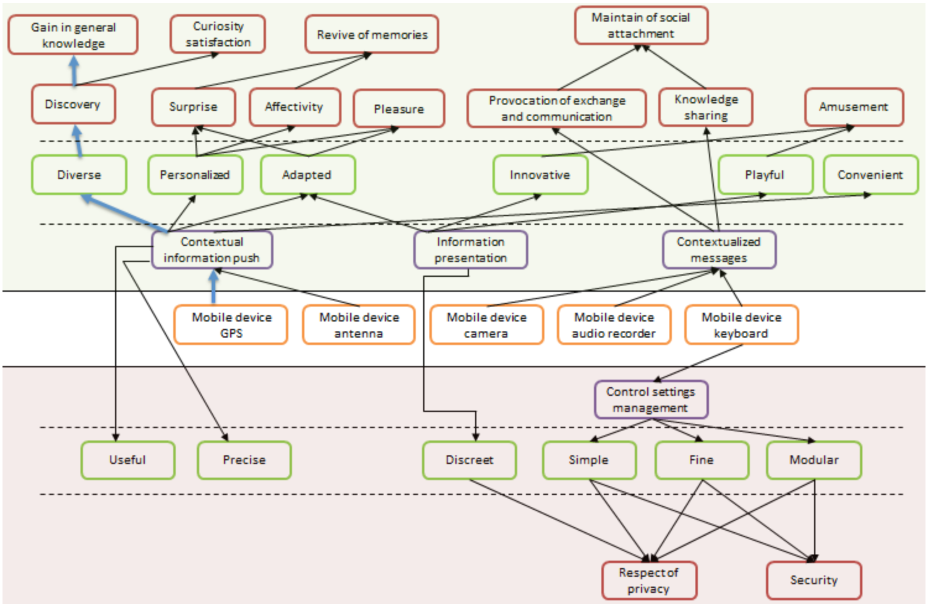


Fig. 2. The WM of Cocoon (partial) – The chain highlighted in blue can be read as follows: the contextual information push feature supported by the GPS of the mobile device presents the quality of being diverse (because serves the user with different types of information, personal and impersonal); this diversity enhances discovery of new things which results in a gain of general knowledge for the user.

- WMs interest

Outcomes from discussions with interactive systems design experts highlighted additional interests of WMs for designing.

– **WMs as support to UI and interaction design**

According to the interaction designer, WMs may support UI and interaction design through qualities. For instance, an expected quality of modularization (see Fig.2) suggests organizing the UI using different blocks.

– **WMs as support to graphic design**

According to the graphic designer, WMs may support his design activity through user-oriented attributes and, especially, through UX elements. More precisely, a WM would inspire him in the choice of the design "ambiance": color, graphical elements, etc. For instance, according to him, he will not hesitate to use joyful colors for Cocoon as the study of worth revealed amusement as an UX element (see Fig.2).

– **WMs as support to system implementation**

According to the software engineer, WMs may provide software developers with directions for coding through functional qualities (e.g. rapidity (response time)).

– **WMs as support to communication**

Interviews were conducted with experts addressing different aspects of design. Yet, WMs appeared to be useful for each of them. All the five participants and, particularly the project manager, agreed that WMs may represent a common support for multidisciplinary teams, giving them an overview of the system regarding different aspects.

WMs as support to implementation was less stressed out by the software engineer than WMs as support to interaction and graphic design. However, the software engineer appreciates that WMs convey "what is important to the user". According to him, as he is not heavily involved in user studies, that "gives meaning to his work".

4 Conclusion and Future Work

This paper presents results of a preliminary work on worth maps in interactive systems design. To our best knowledge, this work represents the first attempt of the WCD framework operationalization by a research team not including the method author.

This paper provides a better understanding of the notion of worth, proposes the PEW framework, and highlights additional interests of WMs for HCI design.

More interestingly, this work highlights numerous directions for future research. In the short term, we will validate the applicability of our framework from its use by other design teams. At the same time, we will conduct further investigations regarding WMs different uses, elicited from this work, for supporting design activities through the development of Colibri, an Advanced Planning System (APS).

In the long term, we will explore the ontology for worth and focus on the development of tools, such as LadderUX[1], for WMs construction.

Acknowledgement. We warmly thank the national SYSEO project (ANR 2010-2014) for its support to this work.

[1] `http://www.ladderux.org/`

References

1. Gutman, J.: A Means-End Chain Model Based on Consumer Categorization Processes. Journal of Marketing 46(2), 60–72 (1982)
2. Reynolds, T.J., Gutman, J.: Laddering Theory, Method, Analysis and Interpretation. Journal of Advertising Research, 11–31 (1988)
3. Logan, R.J.: Behavioral and emotional usability: Thomson Consumer Electronics. In: Usability in Practice, pp. 59–82. Academic Press Professional, Inc. (1994)
4. Thomas, P., Macredie, R.D.: Introduction to the new usability. ACM Trans. Comput.-Hum. Interact. 9(2), 69–73 (2002)
5. Vanden Abeele, V., Zaman, B., De Grooff, D.: User eXperience Laddering with preschoolers: unveiling attributes and benefits of cuddly toy interfaces. Personal Ubiquitous Comput., 451 – 465 (2012)
6. Cockton, G.: From quality in use to value in the world. In: CHI 2004 Extended Abstracts on Human Factors in Computing Systems, pp. 1287–1290. ACM (2004)
7. Cockton, G.: A development framework for value-centred design. In: CHI 2005 Extended Abstracts on Human Factors in Computing Systems, pp. 1292–1295. ACM (2005)
8. Cockton, G.: Designing worth is worth designing. In: Proceedings of the 4th Nordic Conference on Human-Computer Interaction: Changing Roles, pp. 165–174. ACM (2006)
9. Brown, B., Taylor, A.S., Izadi, S., Sellen, A., Kaye, J.J., Eardley, R.: Locating Family Values: A Field Trial of the Whereabouts Clock. In: Krumm, J., Abowd, G.D., Seneviratne, A., Strang, T. (eds.) UbiComp 2007. LNCS, vol. 4717, pp. 354–371. Springer, Heidelberg (2007)
10. Cockton, G., Law, E.C., Hvannberg, E.: Putting Value into E-valu-ation. In: Maturing Usability, pp. 287–317. Springer London (2008)
11. Cockton, G., Kirk, D., Sellen, A., Banks, R.: Evolving and augmenting worth mapping for family archives. In: Proceedings of the 23rd British HCI Group Annual Conference on People and Computers: Celebrating People and Technology, pp. 329–338. British Computer Society (2009)
12. Cockton, G., Kujala, S., Nurkka, P., Hölttä, T.: Supporting Worth Mapping with Sentence Completion. In: Gross, T., Gulliksen, J., Kotzé, P., Oestreicher, L., Palanque, P., Prates, R.O., Winckler, M. (eds.) INTERACT 2009, Part II. LNCS, vol. 5727, pp. 566–581. Springer, Heidelberg (2009)
13. Cockton, G.: Getting there: six meta-principles and interaction design. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 2223—2232. ACM (2009)

The Profile of Law Clerks Using Judiciary Informatics in Turkey*

Zerrin Sungur[1], Özlem Alpu[2], Özlem Oktal[3], and Berna Yazıcı[4]

[1] Anadolu University Faculty of Economics and Administrative Sciences, Department of Labor Economics and Industrial Relations 26470 Eskisehir-Turkey
[2] Eskisehir Osmangazi University Faculty of Arts and Sciences, Department of Statistics 26480 Eskisehir-Turkey
[3] Anadolu University Faculty of Economics and Administrative Sciences, Department of Business Administration 26470 Eskisehir-Turkey
[4] Anadolu University Faculty of Science, Department of Statistics 26470 Eskisehir-Turkey
{zsungur,odogan,bbaloglu}@anadolu.edu.tr,
oalpu@ogu.edu.tr

Abstract. The automation of judicial services got it's start in 1998 in Turkey. In order to increase performance and productivity of these services many regulations have been made in parallel with the innovations and improvements realized in the sector of information technologies. There are many internal users using the judiciary informatics such as chief judges, judges, attorneys general, solicitors, prison officers and law clerks. The aim of this study is to analyze the profile of the law clerks using judiciary informatics in Turkey and evaluate their uses of judiciary informatics in terms of the dimensions of performance and effort expectancy, attitude toward using technology, security and risk and anxiety. Web based questionnaire, which was prepared as a five-point Likert type scale including 4898 law clerks, was analyzed through factor analysis. The empirical research was carried out between July and August 2012 in Turkey.

Keywords: law clerks, e-government, judiciary informatics, Turkey.

1 Introduction

Judicial services have been undergoing a rapid transformation in late 1990s due to the impact of the technological changes. They enabled the delivery of services over the Internet. The automation of judicial services got it's start in 1998 in Turkey. In order to increase performance and productivity of these services many regulations have been made in parallel with the innovations and improvements realized in the sector of information technologies. Information systems that are scalable, multi-user and work-based systems came into question in order to increase performance and productivity of institutions. The automation of judicial services is an important step among the

* This study was supported by Anadolu University Scientific Research Projects Commission under the grant no: 1203E045.

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 675–680, 2013.

e-government practices. Some of the contributions of judiciary informatics can be sited as speed, accuracy, consistency, transparency, effectiveness and efficiency and saving[1].

2 The Job of Law Clerks

Some studies indicate that the early practice was very largely against the emergence of the law clerks as a constantly changing group of inexperienced but bright and well-trained young law-school graduates[2]. The first official reference to the idea of employing assistants for the Supreme Court justices occurred in 1885[3]. The job of law clerk has evolved in name as well as in substance as up from 'secretary', for the 'law clerks', 'law assistant', 'research aide', 'legal assistant' and also 'law examiners'[4].

Knowledge, experience and education are required for a law clerk. He/she has to be knowledgeable about laws, court procedures and government regulations. Also he/she to be knowledgeable about administrative and clerical procedures and systems such as word processing, managing files and records, stenography and transcription, designing forms, and other office procedures and terminology. Knowledge of circuit boards, processors, chips, electronic equipment, and computer hardware and software, including applications and programming are also crucial skills for law clerks.

Job duties of law clerks vary between countries, but some common tasks are;

- preparing affidavits of documents and maintain document files and case correspondence,
- searching for and studying legal documents to investigate facts and law of cases, to determine causes of action and prepare cases,
- researching and analyzing law sources to prepare drafts of beliefs or arguments for review, approval, and use by attorney,
- reviewing and filing pleadings, petitions and other documents relevant to court actions,
- serving copies of pleas to opposing counsel,
- communicating and arbitrating disputes between parties.

3 The Aim of the Study

The aim of this study is to analyze the profile of the law clerks using judiciary informatics in Turkey and evaluate their uses of judiciary informatics in terms of the dimensions of performance and effort expectancy, attitude toward using technology, anxiety, and security and risk.

---

1. `http://www.uyap.gov.tr/tanitim/tarihce.html`
2. Newland, C. A.: Personal Assistants to Supreme Court Justices: The Law Clerks. Oregon Law Review. 40 316 (1961).
3. Mahoney, J. D.: Law Clerks: For better or for Worse? Brooklyn Law Review. 54 323 (1988).
4. Baier, P. R.: The Law Clerks: Profile of an Institution. Vanderbilt Law Review. 26 1130 (1973).

4 Sampling and Obtaining Data

Prepared web based questionnaire was applied to all of internal users including law clerks by The Ministry of Justice General Directorate of Information Technologies and 8840 internal users replied. The population size is 76 592. The internal users, chief judges, judges, attorneys general, solicitors, prison officers and law clerks, were assigned to the questionnaire. The empirical research was carried out between July and August 2012 in Turkey. The sample size of law clerks was 4898. Data were analyzed by SPSS 21.0.

5 Findings and Evaluation

The profile of law clerks based on their demographic characteristics was shown in Table 1. The demographic findings indicate that 61.55 % of law clerks are male and 38.45 % are female respectively. Of the respondents, 16.90 % are between 18-24 years of age, 62.62 % are between 25-34, 16.39 % between 35-44, 4.09 % are 45 years and over.

Table 1. Law clerks' demographic characteristics

Variable	Variable level	Frequency (n= 4898)	%
Gender	Female	1883	38.45
	Male	3015	61.55
Age	18-24	828	16.90
	25-34	3067	62.62
	35-44	803	16.39
	45 years and over	200	4.09
Educational status	Secondary school	8	0.16
	High school	1285	26.23
	MA degree	3466	70.76
	MBA degree	137	2.79
	Doctorate	2	0.04
Work Experience	Less than 1 year	768	15.67
	1-5 years	2103	42.93
	6-10 years	1359	27.74
	11-15 years	323	6.59
	16 years and over	345	7.04
Experience of the use of information system	1-3 years	2488	50.80
	4-6 years	1969	40.20
	7 years and over	441	9.00

Of the sample 0.16 % have graduated from a secondary school, 26.23 % have graduated from a high school, 70.76 % have graduated from a university, 2.83 % have an MBA degree and over. Of the sample, 15.67 % have less than 1 year work experience, 42.93 % have 1-5 years work experience, 27.74 % have 6-10 years work experience and 13.63 % have 11 years and over work experience. Of the respondents 50.80 % have 1-3 years experience of the use of information system, 40.20 % have 4-6 years experience of the use of information system, and 9 % have 7 years and over experience of the use of information system.

Table 2. Results of factor analysis

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.92
Bartlett's Test of Sphericity	Approx. Chi-Square	36371.35
	df	190
	Sig.	0.001

The scale was subjected to principal components analysis with varimax rotation. The KMO values was 0.92 and Barlett's Test of Sphericity reached statistical significance (p< .001), supporting the factorability of the correlation matrix. Principal component analysis revealed the presence of four components, explaining 32.88 per cent, 11.29 per cent, 7.07 per cent and 6.41 per cent of the variance respectively. The responses of law clerks were analyzed by factor analysis and four main dimensions were obtained. These are namely; performance and effort expectancy, attitude toward using technology, anxiety, and security and risk. These four factors that eigenvalues higher than 1 are explaining 57.49 per cent of total variance. Cronbach's alpha coefficient was 0.72 for 20 items in our questionnaire. Statements about variables were prepared in 5 point Likert type scale (1= strongly disagree,…,5= strongly agree). The statements used in the questionnaire were adapted from various studies.[5] These statements are presented in Table 3.

[5] Thomson, R. L., Higgins, C. A., Howell, J. M.: Personal Computing: Toward a Conceptual Model of Utilization. MIS Quarterly. 15 125-- 143 (1991); Davis, F. D.: Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly. 13 319—340 (1989); Compeau, D., Higgins C. A., Huff, S.: Social Cognitive Theory and Individual Reactions to Computing Technology: A Longitudinal Study. MIS Quarterly. 23 145—158 (1999); Moore, G. C., Benbasat, I.: Development of an Instrument to Measure the Perceptions of Adopting an Information Technology Innovation. Information Systems Research. 2 192--222 (1991); Venkatesh, V.: Determinants of Perceived Ease of Use: Integrating Control, Intrinsic Motivation, and Emotion into the Technology Acceptance Model. Information Systems Research. 11 342--365 (2000); Belanger, F., Carter, L.: Trust and Risk in E-government Adoption. Journal of Information Systems. 17 165 --176 (2008); Taylor, S., Todd, P.: Assessing IT Usage: The Role of Prior Experience. MIS Ouarterly. 19 561--570 (1995).

Table 3. Factor loadings, variance percents of factor analysis with varimax rotation

Factor 1: Performance and effort expectancy			
Eigenvalue: 6.58	% of variance: 32.88	cumulative %: 32.88	FA
I would find the system useful in my job on my side.			0.82
Susceptibility to the computer in using information systems make it easier to do my job.			0.76
I would find practical to use information systems.			0.73
I would find beneficial to use information systems in my job on behalf of judiciary systems.			0.65
I would find beneficial to use information systems in my job on behalf of judiciary systems to provide the transparency.			0.62
Factor 2: Attitude toward using technology			
Eigenvalue: 2.26	% of variance: 11.29	cumulative %: 44.18	FA
I am glad of the number of steps in the transaction.			0.71
Information systems help me to make more effective decisions.			0.69
I am glad to be satisfied my information processing requirements.			0.66
I am pleased from the information systems in general.			0.58
The system makes my work more interesting.			0.57
Factor 3: Anxiety			
Eigenvalue: 1.41	% of variance: 7.07	cumulative %: 51.25	FA
I hesitate to use the system for fear of making mistakes I cannot correct.			0.83
It scares me to think that I could lose a lot of information using the system by making wrong transaction.			0.76
I am worried about using the system.			0.72
The system is somewhat intimidating to me.			0.65
It is impossible to correct the mistakes or too difficult.			0.64
Factor 4: Security and risk			
Eigenvalue: 1.25	% of variance: 6.24	cumulative %: 57.49	FA
Knowledge in information system cannot be changed by unauthorized people.			0.73
Unauthorized entry to information system is not possible.			0.72
There is no danger of leaking information from information system.			0.70
Information system is reliable.			0.63
Violations of security in information systems are determined by judicial competent authorities.			0.47

6 Conclusion

In this study, the profile of law clerks using judiciary informatics in Turkey was examined. And also their uses of judiciary informatics in terms of the dimensions of performance and effort expectancy, attitude toward using technology, security and risk and anxiety were analyzed. The demographic findings indicate that it is a job that men prefer more, and 96 % of them under the age of 44. Most of them have graduated from a university. Of the sample 70 % have 1-10 years work experience and half of them have 1-3 years of experience of the use of information systems. The responses of law clerks were analyzed by factor analysis and as a result four main dimensions were namely; performance and effort expectancy, attitude toward using technology, security and risk and anxiety. These four factors that eigenvalues higher than 1 are explaining 57.495 per cent of total variance.

It can be said that law clerks are one of the most active users of judiciary informatics in Turkey. In order to increase performance and productivity in judicial services it seems necessary to know the profile of the internal users especially law clerks in judiciary informatics. In this meaning, the contribution of the findings of this study to HCI can be taken into account.

References

1. Baier, P.R.: The Law Clerks: Profile of an Institution. Vanderbilt Law Review 26, 1125–1177 (1973)
2. Belanger, F., Carter, L.: Trust and Risk in E-government Adoption. Journal of Information Systems 17, 165–176 (2008)
3. Compeau, D., Higgins, C.A., Huff, S.: Social Cognitive Theory and Individual Reactions to Computing Technology: A Longitudinal Study. MIS Quarterly 23, 145–158 (1999)
4. Crump, D.: Law Clerks: Their Roles and Relationships with Their Judges. Judicature 69, 236–240 (1986)
5. Davis, F.D.: Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly 13, 319–340 (1989)
6. Engler, R.: And Justice for All-including the Unpresented Poor: Revisiting the Roles of the Judges, Mediators, and Clerks. Fordham Law Review 67, 1987–2070 (1998-1999)
7. Mahoney, J.D.: Law Clerks: For Better or for Worse? Brooklyn Law Review 54, 321–345 (1988)
8. Moore, G.C., Benbasat, I.: Development of an Instrument to Measure the Perceptions of Adopting an Information Technology Innovation. Information Systems Research 2, 192–222 (1991)
9. Newland, C.A.: Personal Assistants to Supreme Court Justices: The Law Clerks. Oregon Law Review 40, 299–317 (1961)
10. Taylor, S., Todd, P.: Assessing IT Usage: The Role of Prior Experience. MIS Ouarterly 19, 561–570 (1995)
11. Thomson, R.L., Higgins, C.A., Howell, J.M.: Personal Computing: Toward a Conceptual Model of Utilization. MIS Quarterly 15, 125–143 (1991)
12. Venkatesh, V.: Determinants of Perceived Ease of Use: Integrating Control, Intrinsic Motivation, and Emotion into the Technology Acceptance Model. Information Systems Research 11, 342–365 (2000)
13. Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D.: User Acceptance of Information Technology: Toward a Unified View Author(s). MIS Ouarterly 27, 425–478 (2003)

Towards a Communication System for People with Athetoid Cerebral Palsy

Yohan Guerrier[1], Christophe Kolski[1], and Franck Poirier[2]

[1] Univ Lille Nord de France, F-59000 Lille
UVHC, LAMIH, CNRS, UMR 8201, F-59313 Valenciennes
[2] Laboratoire UMR Lab-STICC Université de Bretagne Sud F-56000 Vannes
{yohan.guerrier,christophe.kolski}@univ-valenciennes.fr,
Franck.Poirier@univ-ubs.fr

Abstract. Communication is an important act in the development and empowerment of human beings. Through language, humans communicate their needs, desires, moods... Unfortunately, many physical and mental disabilities deprive some people of such communication means. Nowadays various Augmentative and Alternative Communication (AAC) systems exist in order to help people with disabilities. Virtual keyboards are the most common AAC systems for physical disabilities. Concerning mental disability, there are tools based on pictograms. This paper is divided into two parts. First, we put forward a critical review of various AAC systems with a focus on users with athetoid cerebral palsy. Second, the paper presents work in progress concerning a communication system for such users.

Keywords: Communication, mobility, cerebral palsy (CP), slurred speech, communication aid, COMMOB.

1 Introduction

« *Communication is an act so natural that many people forget its importance. Without communication, we are lonely beings in a completely disordered world. Emotions cannot be expressed and therefore there would be no relationship between humans* ». At first sight, this observation may seem totally impractical and straight out of a science fiction book. In reality, the statement came from an anonymous person with cerebral palsy and in so doing speaks for a large category of the population living in a world apart. It underscores the fact that for many people with intellectual or physical disabilities, it is extremely challenging to communicate verbally – but also often through a system [15, 16] – for example, when having to formulate correct sentences. The challenge is even greater for those affected with athetoid cerebral palsy [9]. In addition to experiencing major difficulties in making gestures (e.g. inability to use a keyboard), these people find it difficult to pronounce words when producing correct sentences. This is because the verbal utterances are deformed due to dysarthria [14]. Many people want to express a lot of things, but they are unable to do so. People with athetoid cerebral palsy (CP) also have problems with manual gestures due to

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 681–688, 2013.

involuntary movements due to neurons damaged from lack of oxygen at birth. These neurons continue to send electrical impulses to the muscles to dysfunctional arms and legs, thus causing involuntary movements with high amplitudes [9]. In this context, we propose the COMMOB system software to facilitate everyday communicational activities with systems and other people. The system cannot be directly transferred to other disabilities because of the specific needs of athetoid CP users.

The article begins with a state of the art about different categories of communication systems for people with cerebral palsy. Different communication needs of such people (considered in this paper as users) are accordingly presented. Then, our current work on the COMMOB communication system is explained. The article concludes with a discussion and research perspectives.

2 Related Work

In [7], a critical study of various means of communication, more or less easily exploitable by users with athetoid CP, was conducted. It identified four categories of communication aids: virtual keyboards, physical aids, speech recognition system, brain-computer interface. (Please note that optical motion tracking is not considered in this study, because it is not possible due to the uncontrolled movements of the people with cerebral palsy). In what follows, we briefly present some virtual keyboards and assistive communication systems that seem potentially interesting for users with cerebral palsy. These systems were chosen in relation to their level of use in the field of disabilities. The first category of communication aids are virtual keyboards such as:

- **K-Thot optimized** [1]: This has four letters per key. The letters are grouped by occurrence frequency in French. For example, the letter "q" is on the same key as the letter "u". This process reduces the distance traveled by the pointer (mouse), and consequently, it reduces physical effort and reduces user fatigue.
- **HandiGlyph** [3]: This aid has three ambiguous keys and a command key. A regular scanning focus rotates past the four keys. Each ambiguous key corresponds to a family of letters that share a similarity of form (diagonal stroke, curve, straight line). The expected word is deduced by a linguistic predictor like for all the ambiguous keyboards.
- **KEYGLASS** (http://keyglass.free.fr/): This system proposes the most probable letters on four semi-transparent keys around the last key pressed. For example, if the user presses the letter "b", letters "e, a, u, o" will appear around the "b" key. This process reduces the distance traveled by the mouse and therefore makes text input faster.
- **Chewing Word** [5]: This keyboard combines several methods to accelerate the input. First, moving the most probable letters closer to the last key pressed. Secondly, the predicted word appears in a bubble just above the last key pressed.
- **Sibylle** [10]: The system consists of a virtual keyboard comprising a set of sub-keypads that allows the entering of characters or full words. The cursor successively highlights each key of the active keypad that can then be selected by a

single-switch selection process. It also has sophisticated letter and word prediction components which dynamically calculate the most appropriate letters or words for a given context. This method can reduce more than 50% the number of key presses.

- **UKO-II** (http://wiki.cogain.info/index.php/UKO-II): This contains all the letters on only 4 keys. To enter a letter, the user indicates the number of the key where the desired letter is situated. Then it enters the letter rank in the selected key. This process reduces moving the mouse over a large number of keys.
- **K-hermes** [6]: This is an ambiguous multitap keyboard with several letters per key. The user presses the desired key n times to select the letter of rank n. For example, if the first key is for letters "a", "b", "c", to enter the letter "b", the user must press the key twice. In addition, in order to speed up the input, the keyboard offers word prediction.

A lot of work on virtual keyboards for mobile devices is also available in [11]. In general, virtual keyboards allow you to enter text, while requiring a significant physical effort over a long period, despite the many optimizations features described above. To summarize, virtual keyboards are not well equipped to enter a long text for people with athetoid CP. To improve virtual keyboards, the predictions of words need to be more efficient by reducing the number of movements during input.

The second category concerns physical communication aids:

- **Joystick EdgeWrite** [17]: This system allows text input through the movements of a power wheelchair joystick. The user forms the letter by moving the joystick along the physical edges and into the corners of a square bounding the input area.
- **Entering row / column with two buttons** [2]: This virtual keyboard has a cursor that highlights a key at the crossing of a line and a column. The user has two physical buttons, one for changing the line, the other for changing the column. After a certain period, the highlighted key is selected.
- **Guide finger** [8]: This consists of a plastic plate with holes to access the keys on a physical keyboard. This system avoids pressing multiple keys at the same time.

These physical aids have the drawback that they must be used since childhood to be very effective. If the user has not used this kind of equipment when his/her brain still had a significant plasticity, it can be difficult, if not impossible, to use the Guide finger as an adult [4].

The third category includes speech recognition systems, where a voice recognition approach attempts to adapt recognition systems for people with dysarthria [14]. These systems seek to standardize as much as possible the voice of disabled people. However, the error rate remains sufficiently significant for voice recognition not to be considered as very efficient.

The fourth category concerns Brain-Computer Interfaces (BCI) [18]: such interfaces aim at controlling a computer by using brain activities detected by electrodes placed in a helmet. This allows to single actions only (such as a keypress) by means of a scanning system. It requires high concentration [12] and is not usable for the moment on a mobile device because it requires too much resources and heavy equipment. Moreover, this system has a high error rate (80% [12]).

In short, these different categories of AAC systems are not satisfactory for users with athetoid CP. A new type of system is thus envisaged in the section that follows.

3 Needs Analysis and Specification of the COMMOB System

The name of system is a combination of two words: *COMmunication* and *MOBility* because COMMOB is an assistive communication system for use at home or on the go. The well-known concept of pictogram is at the basis of this communication system. Pictograms are used to accelerate the production of sentences, by proposing, whenever possible, a set of context sensitive pictograms. For example, if the user is in a railway station, the program offers symbols only related to the location, or if the user selects the symbol "I would like to drink", the program offers only symbols denoting drinks. The user can also use a predictive virtual keyboard (see section 2) to enter words not represented by pictograms (for example names of railway stations). The keyboard is also context sensitive. It can propose words according to the interaction context that can be characterized through various sensors of the system. For example, the location of the user will be retrieved by the GPS chip, the sound volume will be recovered by the microphone, ...

The complexity of the problem lies in the possibility of access and to express a maximum of words or sentences through the pictograms for the user to enter minimum words. There is a strong link between the virtual keyboard and the module managing the pictograms. A set of predefined dialogues is proposed so that the user does not have to enter all the words. The user also has the possibility to prepare future dialogue in advance before using the system. These two features are intended to speed up text input, so to make the dialogue more fluid. Other modules may be added in order to extend the basic functionalities. For example, a speech synthesizer to read sentences produced. In the case of a CP user who is a computer scientist, a module exists to help him or her program by accelerating the entry code. The system provides the user with a set of software components (bricks), each specialized with a specific feature (such as GUI design) and represented by pictograms so that he or she can assemble the software.

Figure 1 represents the needs of users with athetoid CP concerning oral and written communication. Each module of COMMOB meets a specific need of the disabled person. These needs are grouped into themes (for example, "On the street" regroups communication needs of a disabled person in such a context). Obviously, some needs may be present several times in different bubbles. However, we chose to put only the most representative bubble so as not to overload Figure 1.

The COMMOB system is currently in a prototyping phase. It is written in Java under MS Windows 7 that allows an easy finding of already developed modules which can be reused or adapted. The tablet used is a HP Slate. It provides an 8.9 inch touchscreen (23 cm). The size of the tablet computer allows its installation on a

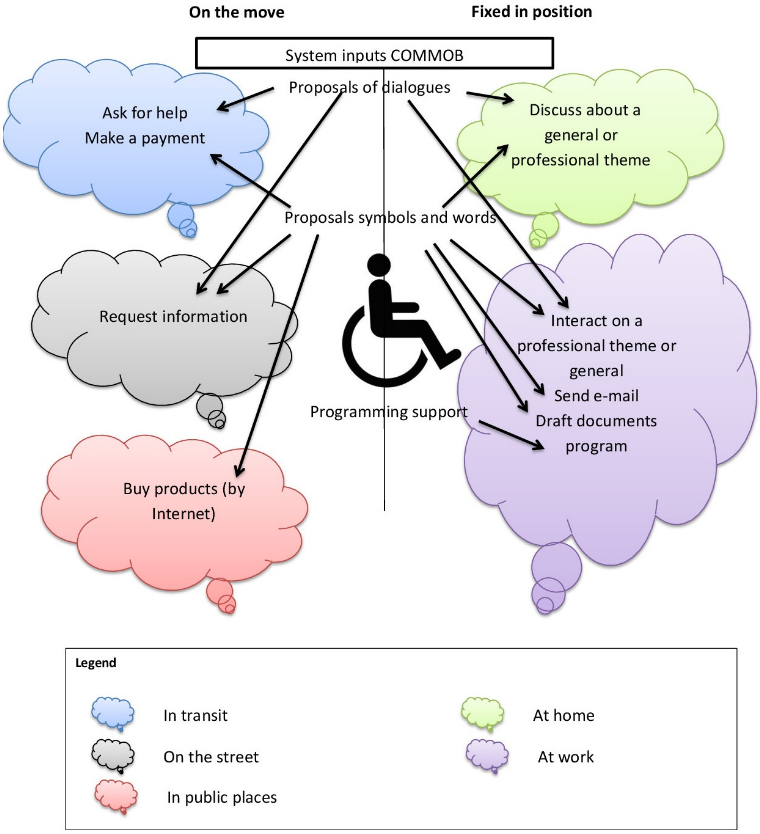


Fig. 1. Potential contributions of COMMOB

wheelchair to be used everywhere. It has a USB port that connects an infrared system to control the cursor with a joystick installed on the wheelchair.

Figure 2 shows an example of formulating sentences from pictograms associated with words; for complementary text entry, the use of a virtual keyboard is necessary. Indeed, some words cannot be represented by pictograms, such as the names of railway stations. The user must complete the dots using the virtual keyboard, in this case clavicom NG (http://code.google.com/p/clavicom/). This keyboard provides the user with a set of predicted words to accelerate the entry.

The device, based on joystick and touch pad, is embarked directly on the wheelchair. Figure 3 illustrates the following situations: in an office (a), in a public place (b), outside (c). The user moves the mouse pointer with the wheelchair joystick. The connection is made through the electronic unit, which is located above the left handle (c). Information flows in the infrared beam. The validation is done through a button next to the joystick (d).

This method (pictograms + virtual keyboard) allows the user to enter words that cannot be readily represented by pictogram(s).

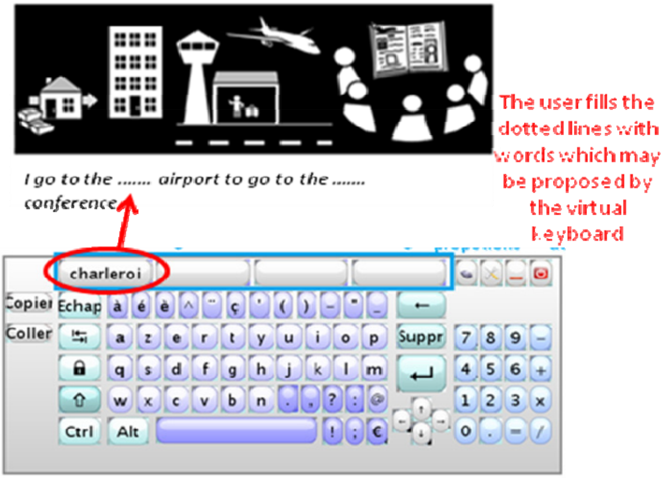


Fig. 2. Example of formulation of a sentence from words and pictograms

Fig. 3. Embarked COMMOB system, used in different situations

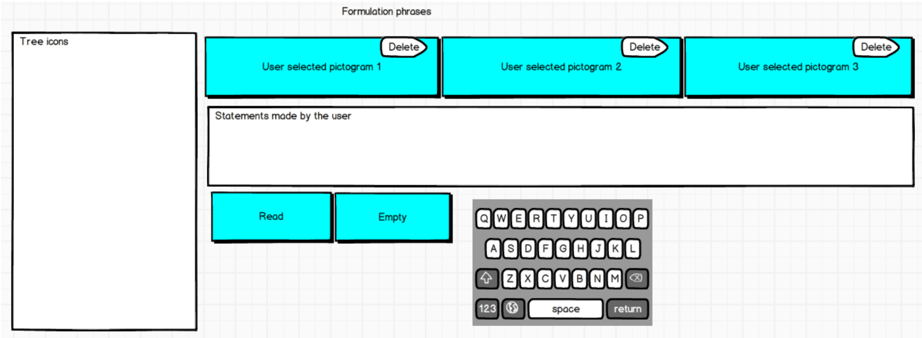


Fig. 4. Screenshot of an example of sentence generation

Figure 4 shows a screenshot of the GUI module for formulating sentences with pictograms. The user selects symbols from the tree to the right of the Figure.

4 Discussion and Future Work

In this paper, AAC systems for users with athetoid CP have been envisaged. These communication systems are grouped in different categories which all have their advantages and disadvantages. We believe that in mobile contexts, the virtual keyboard (available and controllable with a joystick installed on a wheelchair) is a tool well suited for text input. Accordingly, our future perspectives take the form of an integrated communication support within the scope of the doctoral thesis of the first author. It aims to help people with athetoid CP who have difficulty communicating with systems and humans in various situations (at home, in mobility...). To do this, we chosen to use pictograms to accelerate the input of sentences, which are supplemented by a virtual keyboard (for the names of cities, for example).

To sum up, the main features of our system are: the use of (1) pictograms in order to speed up the text input; and (2) context-sensitive prediction. A prototype is under development (the system is currently composed of different mock-ups). Our aim is to finish a functional prototype and to conduct user testing as soon as possible. These tests will be based on a scenario with several activities (on the move or not, using pictograms or not). Results will be analyzed in terms of error rate, number of pictograms selected, number of entered words, using standard metrics of the domain (KSPC, ksr ...) [11].

Acknowledgments . This research work has been supported by CISIT, the Nord-Pas-de-Calais Region, the European Community (FEDER), the ANR (TecSan 2011 SI-FaDo project n° ANR-11-TECS- 0014). The authors thank the anonymous reviewers and Michel Labour for their numerous remarks.

References

1. Baas, M., Guerrier, Y., Kolski, C., Poirier, F.: Système de saisie de texte visant à réduire l'effort des utilisateurs à handicap moteur. In: Proceedings of the Ergonomie et Informatique Avancée Conference, October 12-15, pp. 19–26. ACM Press, Biarritz (2010)
2. Baljko, B., Tam, A.: Indirect Text Entry Using One or Two Keys. In: Proc. ASSETS 2006, Portland, Oregon, USA, October 22-25. ACM Press (2006)
3. Belatar, M., Poirier, F.: Text entry for mobile devices and users with severe motor impairments: handiglyph, a primitive shapes based onscreen keyboard. In: Proc. ACM ASSETS, Halifax, Canada, pp. 209–216 (October)
4. Kolb, B., Gibb, R.: Brain Plasticity and Behaviour in the Developing Brain. J. Can. Acad. Child Adolesc. Psychiatry 20(4), 265–276 (2011)
5. Grange, A.: L'interface du clavier virtuel Chewing Word. In: Actes d'IHM 2010, 22e Conférence Francophone sur l'Interaction Homme-Machine, pp. 237–240. ACM Press (September 2010)
6. Guerrier, Y., Baas, M., Kolski, C., Poirier, F.: Comparative study between AZERTY-type and K-Hermes virtual keyboards dedicated to users with cerebral palsy. In: Stephanidis, C. (ed.) Universal Access in HCI, Part II, HCII 2011. LNCS, vol. 6766, pp. 310–319. Springer, Heidelberg (2011)
7. Guerrier, Y., Kolski, C., Poirier, F.: Comment communiquer lorsqu'on est un utilisateur infirme moteur cérébral (IMC) de type athétosique? Etude de cas des différents moyens de communication. In: Proc. of the ERGO-IHM Conference, October 17-19, pp. 51–54. ACM, Digital Library, Biarritz (2012)
8. `http://djtechtools.com/2012/08/19/momentary-fx-in-ableton-live-advanced-midi-mappings/`
9. `http://cerebralpalsy.org/about-cerebral-palsy/what-is-cerebral-palsy/`
10. `http://www.info.univ-tours.fr/%7Eantoine/SIBYLLE/Sibylle_en/index.html`
11. MacKenzie, I.S., Soukoreff, R.W.: Text entry for mobile computing: Models and methods, theory and practice. Human-Computer Interaction 17, 147–198 (2002)
12. Manyakov, N.V., Chumerin, N., Combaz, A., Van Hulle, M.M.: Comparison of Classification Methods for P300 Brain-Computer Interface on Disabled Subjects. Computational Intelligence and Neuroscience 2011, Article ID 519868, 12 pages (2011)
13. Miller, F.: Cerebral palsy. Springer (2009)
14. O'Sullivan, S.B., Schmitz, T.J.: Physical rehabilitation, 5th edn. F. A. Davis Company, Philadelphia (2007)
15. Pereira, J.: Handbook of Research on Personal Autonomy Technologies and Disability Informatics. IGI Global (2010)
16. Waller, A.: Interpersonal Communication. In: Stephanidis, C. (ed.) The Universal Access Handbook, pp. 47-1–47-12. CRC Press Taylor & Francis Group (2009)
17. Wobbrock, J.O., Myers, B.A., Aung, H.H., LoPresti, H.F.: Text Entry from Power Wheelchairs: EdgeWrite for Joysticks and Touchpads. In: ASSETS 2004, Atlanta, GA, USA, pp. 110–117. ACM Press (2004)
18. Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G., Vaughan, T.M.: Brain–computer interfaces for communication and control. Clinical Neurophysiology 113, 767–791 (2002)

Towards Supporting the Existing Workplace Practices of a Community of Brazilian Healthcare Professionals

Roberto Calderon[1], Sidney Fels[1], and Junia Anacleto[2]

[1] The University of British Columbia
[2] Universidade Federal de São Carlos
roberto@alumni.ubc.ca, ssfels@ece.ubc.ca, junia@dc.ufscar.bc

Abstract. With the increasing affordability of computers, displays and telecommunications, the scenario of introducing digital Information and Communication Technologies (ICT) into communities with little or no previous exposure to computing has become common place. Understanding how ICT affects the functioning of such communities is important for determining design and introduction strategies that can minimize the disruption of well established practices in said scenarios. We designed and introduced a ticketing system within a community of Brazilian healthcare professionals that have little or no previous exposure to computing. Visualizing individual people led to tasks directed towards particular individuals. Visualizing people interactions promoted open-ended and communal tasks. We observed that professionals circumvented the original design of the system to introduce unimplemented functionalities and support their well-established social-based information management practices.

Keywords: Visualization, Situated Displays, Healthcare, Collaboration, Social Capital.

1 Introduction

One of challenges regarding the introduction of Information Communication Technologies (ICT) into the workplace is understanding the effects of such technologies on the functioning of the communities that work there. In the present paper, we investigate the challenges of supporting the existing workplace practices within a community of healthcare professionals in a Brazilian centre for mental care. In particular, we aim at documenting how a community that has had little or no previous exposure to digital Information and Communication Technologies adopts the introduction of digital ICT and how the functioning of such community is affected by the new technology.

Through a previous ethnographic study [2] of this community, we uncovered a communication system based on verbal communication and pen-and-paper records

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 689–696, 2013.

that is sustained by informal socialization carried within the workplace throughout the day. Interpersonal communication and impromptu gatherings allow for efficient flow of information, that allows for flexible hierarchical structures and efficient task and patient status tracking.

Previous research has uncovered that situated public displays [10], and in particular the representation of ad-hoc work-flows [5,13] can promote communication and enhance collaboration between healthcare professionals. However, the appropriation patterns and how such technology affects the social structure supporting collaboration within communities with low technical background is still unexplored. The present research focuses on determining the characteristics of ICT that could support the ad-hoc socialization practices that are essential to the community being studied.

Using a user-centred design we designed a ticketing system utilizing a situated public display and compared two types of visualizations: one depicting individual actions and a second one representing collaborative and group actions. We observed that the community of healthcare professionals modified the original purpose of the technology in order to accommodate unimplemented functionalities (e.g. capitalizing co-workers' names and adding task dates), and to fit their long-established information and communication practices (e.g. using the ticketing system to engage in conversation, and to publicize messages of sympathy and support towards co-workers).

2 Related Work

Systems to provide social awareness in order to promote collaboration within a workplace include the Notification Collage [6], and IM here [8]. These projects build upon a long history of research [4, 11] on the effect of public displays to promote collaboration within co-workers. Yet, this research often takes place within an office or research lab environment, which have different social dynamics and work-flow patterns from their healthcare counterparts. The benefits of public displays to promote collaboration in healthcare settings have been previously identified by Lasome et al. [10], who argue that public displays, like task boards, are naturally used to provide increased task management and shared awareness leading to more efficient problem solving. Within a healthcare setting, some researchers [5,13] have investigated the use of task management systems to promote collaboration. This body of work suggests that the introduction of technology within the healthcare workplace changes how information is shared within members of the workplace, often modifying how interpersonal connections are made. Such investigations argue that the introduction of technology, particularly public displays to track and publicize workplace tasks, has many benefits such as an increase on collaboration [1] and socialization [9] patterns. The need to document the differences between HCI in the developed world and HCI for the developing one have been previously stressed in the literature [3]. Similarly, the need to investigate the role of ICTs for healthcare communities with little exposure to technology has been previously stressed in the literature [7].

3 CAIS Hospital for Mental Care

CAIS Clemente Ferreira (the hospital) located in Lins, São Paulo, Brazil, is a non-profit state-run institution recognized as a Brazilian centre of excellence in mental health focusing on psychiatric disorders and neurological damages. The hospital currently employs over 380 professionals and provides care to more than 800 patients into eight care units serving different types of mental disorders and age groups. The architectural setting of the hospital - isolated urban location, corridors exceeding 100 meters in length and thick walls built in the early 1900s - make the installation of information and communication technologies a challenge, that paired with low governmental budgets have fostered a technology-free environment. In addition, many of the healthcare professionals have not been directly exposed to computing in their daily lives, as personal computers are rare and expensive in the area. As a result, healthcare professionals working at the hospital have developed, over the course of 25 years, an informational system based on pen-and-paper and verbal communication. The community studied is formed of auxiliary nurses, registered nurses, doctors, physiologists, phonologists and administration personnel. Following an ethnographic study we discovered that this community leverages socializing practices (e.g. storytelling) to communicate and collaborate.

3.1 Social Work-Flows Currently Adopted by the Community

In order to identify the natural practices adopted by the community being observed, we performed an ethnographic study where we *shadowed* members of the healthcare community [2]. In this study we identified the existence of informal gatherings and the use of storytelling as the core factors supporting collaboration between members of this community. Informal gatherings taking place on corridors, nursing rooms and hidden corners serve to promote verbal communication where tasks are informally delegated or tracked through playful conversations. This promotes an on going process of storytelling involving both personal and work-related stories that serve to efficiently track the status of important tasks, e.g. a new patient arrival. In this process of storytelling, professionals of all ranks are socially expected to participate and engage into vivid and friendly conversations. As a result, the otherwise inflexible job hierarchies and descriptions become flexible. That is, often people go off their job description and daily tasks to resolve issues discussed during these social meetings. We found that these naturally adopted collaboration practices are enhanced by a culturally rooted understanding of community where verbal and playful sociability is celebrated and exercised on every interpersonal encounter.

4 Design of a Task Managing System to Support Informal Gatherings and Storytelling

Our ethnographic study uncovered that daily tasks within the hospital are managed through informal gatherings and storytelling. In order to understand how ICTs can support these practices we designed a task managing application and compared two visualizations. A task managing kiosk, allowing healthcare professionals to manage tasks using a keyboard and mouse (Fig. 1), was designed following a user-centric iterative process. Two graphical representations of were made public for one month each, in order to measure the effects of different types of data visualization on the community. The first visualization (individual-centric) was designed to represent the individual count of resolved tasks. The second visualization (community-centric) was designed to represent individual actions in relationship to the community.

The first visualization (Fig. 2-A) consisted of circles representing each of the professionals of the wing. The radius of each circle was computed as the number of issues resolved by a person R_i. The positioning of each circle was randomized at screen initialization. Understanding this visualization was straightforward, as people resolved more tasks, a circle with their name increased in size.

The second visualization (Fig. 2-B) focused on representing the relationship of each person to other members of the group and the group as a whole. Each person was represented as a circle whose radius was computed as the ratio of issues resolved by a person R_i to all issues resolved by all the members of the group R, i.e. R_i/R. Circles were then arranged equidistantly from each other in a circle, and lines were drawn to represent links created between the person that created an issue and the person that resolved it. The width of such lines was computed as the ratio of issues resolved by a person i and issued by person j , $R_{i\to j}$, to all issues resolved by person i, that is $R_{i\to j}/R_i$.

The size of circles represented people's *ranking*, rather than their number of tasks resolved. Therefore, changes in size would only be apparent if people solved considerably more tasks than their similarly-ranked peers. Additionally, thicker lines often meant that people would interact between each other more often. To make this clearer to the users, a legend was included on the bottom right part of the visualization explaining that larger circles "resolved more" and smaller circles "resolved less", while thinner lines depicted "less cooperation" and thicker lines represented "more cooperation".

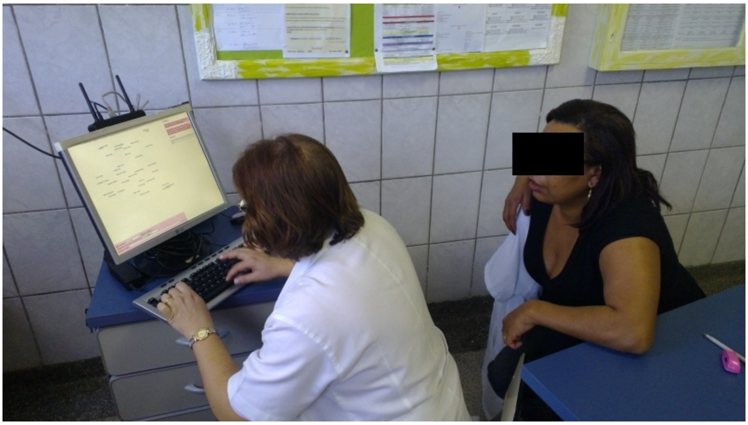


Fig. 1. Healthcare professionals use the public task-managing kiosk

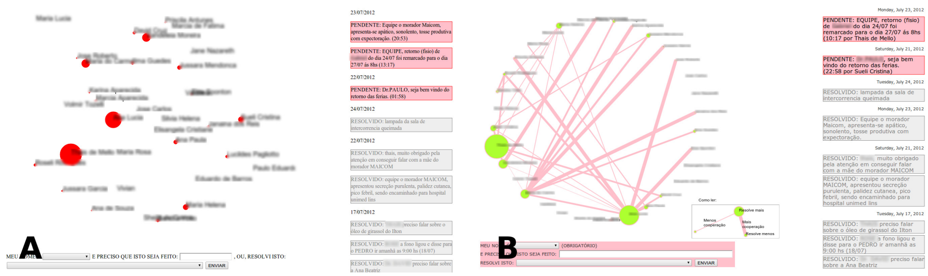


Fig. 2. Left: Individual-centric visualization showing people as isolated circles. (Screen-shot taken with all tasks collected during all 8 weeks.). Right: Community-centric visualization showing circles whose radius and connections depict people in relationship to the community or other people. (Screen-shot taken with all tasks collected during all 8 weeks.).

5 Methodology Used for Semantic Analysis of Tasks

In order to understand the motivations driving the usage of the kiosk we performed a semantic analysis of all tasks submitted through our prototype. Our analysis was theoretically grounded on the Interactive Places Framework proposed by Memarovic et al. [12]. The Interactive Places Framework proposes understanding content produced and consumed through public displays according to three strands: information production, means of disseminating such information and awareness of such information. Memarovic et al.'s conceptual framework models public display content as information created and consumed by either individuals or groups of people (i.e. with different levels of *exclusivity*) that when made available through the public display can be either explicit or implicit to a particular group of people or an individual (i.e. promoting different types of *awareness*).

5.1 Exclusivity

When content is directed towards anybody that comes in contact with the public display said content is said to be *inclusive*, however, when it is limited to a particular individual or group is said to be *exclusive*. When semantically analysing tasks submitted to our ticketing system we characterized *exclusive* content as tasks directed towards a particular healthcare professional or a group (e.g. “[MARTHA]: please fix the lamp on the bathroom” or “NIGHT TEAM: don't forget to stamp your forms after daily check-ups.”). We characterized *inclusive* content as generalized tasks not directed towards a particular person or a group of people, but to the whole community of healthcare professionals (e.g. “Patient [X] will not be receiving visitors this weekend.”).

5.2 Awareness

Although a public display might be broadcasting information in a public manner not all people that come in contact with the content are able to make sense of it, or find it relevant to their interests and values. *Explicit* content portrays information about the interests and values of particular individuals or groups of people within the community. *Implicit* content is generalized information not pertaining a particular individual, or interest within the community. When semantically analysing tasks we considered *explicit* messages to be exemplified by calls for action from either particular individuals or the community (e.g. "[MARTHA]: please fix the lamp on the bathroom" or "TEAM don't forget to stamp your reports"). By contrast, tasks conveying generalized states of the workplace or unaddressed messages of conviviality were considered as *implicit* (e.g. "TEAM Have a good weekend" or "Emergency room light stopped working").

6 Findings

We observed that professionals modified the original intended usage of the system to accommodate their existing practices. The original design of the ticketing system did not allow assigning tasks to particular individuals or groups. However, shortly after deployment some users begun addressing individuals by commencing new tasks with capitalized names (e.g. "[MARTHA] please fix..."). By the third week, this was considered a regular way to begin a message addressed to a particular individual or a group of people. Similarly, professionals created a way to determine task timings and deadlines by adding a date at the end of the submitted tasks. (e.g. "[MARTHA] attend Occupational Health day 15/7 8:30"). In both scenarios repetitive usage led to a common practice adopted by all members of the community of healthcare professionals.

We found that although the system was originally designed as a ticketing system to publicize tasks regarding the workplace healthcare professionals used it to post non-task messages to other professionals (e.g. "[Doctor X] welcome back from your vacation"), encouraging messages (e.g. "May we find strength to do our jobs today") and open ended socialization phrases (e.g. "Good day" or "Good weekend everyone"). In some cases we found traces of conversational posts (e.g. "I thank everyone for the support and help. May god bless you. Kisses." to which somebody replied "We are with you [Martha]").

6.1 Semantic Analysis Findings

We observed that during the individual-centric visualization the majority of tasks submitted were often explicitly directed towards a particular goal, person, or group of people (e.g. "Please stamp the examination request from [JACK]..." and "[MARTHA] please take care of my baby"). The majority of tasks submitted during this visualization often called for specific actions from members of the community.

By contrast, during the deployment of the community-centric visualization the number of implicit tasks depicting a status of the workplace (e.g. "[MARTHA], [GUS]'s wheel chair is wet") or messages to promote general awareness of a future or past event (e.g. "TEAM... [JACK] will go on the outing to the zoo day 04.12.2012...") greatly increased.

We observed that during the individual-centric visualization the majority of task were exclusive, addressing particular healthcare professionals or groups of people, and often contained the capitalized name of the addressee in the body of the task. However, when the community-centric visualization was deployed the number of inclusive tasks, i.e. tasks not addressed to a particular individual and concerning anybody that came in contact with the display, surpassed that of exclusive, i.e. directed, tasks.

7 Conclusion and Future Work

The present study investigates how ICTs can help support the existing practices of a community of healthcare professionals in a Brazilian centre for mental care. We performed an ethnographic investigation of this community and uncovered that informal gatherings and storytelling are essential to the centre. These practices promote efficient task delegation and tracking, and a sense of community that has as benefit a better functioning workplace.

A ticketing system using a public kiosk was designed using a user-centred design process to support task tracking and two graphic representations of the data collected were compared. We found that a visualization depicting individual actions promoted tasks demanding specific actions from particular individuals or groups of people. By contrast, we found that a community-oriented visualization enticed healthcare professionals to publicize tasks that were open-ended and usable by all members of the community using the system.

Our future work aims at understanding the role that community reciprocity and common interests play in adopting digital information and communication technologies. We are applying the lessons learned in the above study to design a system to support open-ended and social notices that can be used by all the community.

Acknowledgements. We thank our funding agencies: GRAND, FAPESP, CAPES/DFAIT for their support. We thank the CAIS Hospital staff, MAGIC and our colleague, Jônatas Leite de Oliveira, for his help and insight with this project. The research was performed under the CAIS Ethical Statement.

References

1. Bardram, J.E., Hansen, T.R., Soegaard, M.: Awaremedia: a shared interactive display supporting social, temporal, and spatial awareness in surgery. In: Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work, CSCW 2006, pp. 109–118. ACM, New York (2006)
2. Calderon, R., Fels, S., Oliveira, J.L., Anacleto, J.: Understanding NUI-supported Nomadic Social Places in a Brazilian healthcare Facility. In: IHC 2012 (2012)
3. Chetty, M., Grinter, R.E.: HCI4D: hci challenges in the global south. In: CHI EA 2007, pp. 2327–2332. ACM, New York (2007)

4. Fitzpatrick, G., Mansfield, T., Kaplan, S., Arnold, D., Phelps, T., Segall, B.: Augmenting the workaday world with elvin. In: Proceedings of the Sixth European Conference on Computer Supported Cooperative Work, pp. 431–450. Kluwer Academic Publishers, Norwell (1999)
5. Fouzi, L., Zidani, A., Chorfi, A.: A web application for supporting healthcare tasks with a groupware planning approach. In: 2012 International Conference on Information Technology and e-Services (ICITeS), pp. 1–6 (March 2012)
6. Greenberg, S., Rounding, M.: The notification collage: posting information to public and personal displays. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2001, pp. 514–521. ACM, New York (2001)
7. Ho, M.R.: Communication and computing in health facilities of southwest Uganda. In: CHI EA 2010, pp. 4207–4212. ACM, New York (2010)
8. Huang, E.M., Russell, D.M., Sue, A.E.: Im here: public instant messaging on large, shared displays for workgroup interactions. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2004, pp. 279–286. ACM, New York (2004)
9. Kim, J.-T., Lee, J.-H., Lee, H.-K., Paik, E.-H.: Design and implementation of the location-based personalized social media service. In: ICIW 2010, May 9-15, pp. 116–121 (2010)
10. Lasome, C., Xiao, Y.: Large public display boards: a case study of an OR board and design implications. In: Proceedings of the AMIA Symposium (January 2001)
11. MacIntyre, B., Mynatt, E.D., Voida, S., Hansen, K.M., Tullio, J., Corso, G.M.: Support for multitasking and background awareness using interactive peripheral displays. In: Proceedings of the 14th Annual ACM Symposium on User Interface Software and Technology, UIST 2001, pp. 41–50. ACM, New York (2001)
12. Memarovic, N., Langheinrich, M., Alt, F.: The interacting places framework: conceptualizing public display applications that promote community interaction and place awareness. In: Proceedings of the 2012 International Symposium on Pervasive Displays, PerDis 2012, pp. 7:1–7:6. ACM, New York (2012)
13. Wilson, S., Galliers, J., Fone, J.: Not all sharing is equal: the impact of a large display on small group collaborative work. In: Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work, CSCW 2006, pp. 25–28. ACM, New York (2006)

Traveller: An Interactive Cultural Training System Controlled by User-Defined Body Gestures

Felix Kistler[1], Elisabeth André[1], Samuel Mascarenhas[2], André Silva[2], Ana Paiva[2], Nick Degens[3], Gert Jan Hofstede[3], Eva Krumhuber[4], Arvid Kappas[4], and Ruth Aylett[5]

[1] Human Centered Multimedia, Augsburg University,
Universitätsstr. 6a, 86159 Augsburg, Germany
[2] INESC-ID / Instituto Superior Técnico,
Av. Prof. Dr. Cavaco Silva, 2744-016 Porto Salvo, Portugal
[3] Wageningen University,
Hollandseweg 1, 6706KN Wageningen, Netherlands
[4] Jacobs University,
Campus Ring 1, 28759 Bremen, Germany
[5] Heriot-Watt University,
EH14 1AS, Edinburgh, United Kingdom
{kistler,andre}@informatik.uni-augsburg.de,
{samuel.mascarenhas,andre.silva,ana.paiva}@gaips.inesc-id.pt,
{nick.degens,gertjan.hofstede}@wur.nl,
{e.krumhuber,a.kappas}@jacobs-university.de,
r.s.aylett@hw.ac.uk

Abstract. In this paper, we describe a cultural training system based on an interactive storytelling approach and a culturally-adaptive agent architecture, for which a user-defined gesture set was created. 251 full body gestures by 22 users were analyzed to find intuitive gestures for the in-game actions in our system. After the analysis we integrated the gestures in our application using our framework for full body gesture recognition. We further integrated a second interaction type which applies a graphical interface controlled with freehand swiping gestures.

Keywords: User Defined Gestures, Kinect, Full Body Tracking, Depth Sensor, Interaction, Interactive Storytelling, Cultural Training.

1 Introduction

Experience-based role play with virtual agents offers great promise for social training. In this paper, we present a system called Traveller (Train for Virtually Every Locality) that makes use of such an approach to educate young adults (18-25) in cultural sensitivity. The system implements a virtual storytelling environment in which the users learn by finding out how to appropriately interact with characters simulating different cultures as defined by Hofstede [5].

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 697–704, 2013.

It is vital to the success of experience-based learning with virtual characters that the user is able to interact with them in a socially believable manner. As depth cameras have become broadly available with the Microsoft Kinect[1], we decided to make use of novel full body interaction techniques that allow trainees to practice culturally-varying non-verbal behaviors directly. However, the identification of intuitive gestures that are expressive enough to enable meaningful interaction with virtual characters is a challenge for the interaction designer. Given that gesture sets are usually chosen by the developers themselves, they do not necessarily have to be intuitive for the majority of users.

An approach that employs a user-defined gesture set has been presented by Wobbrock et al. for surface computing [14], and was adapted for other areas, such as public displays [9] or human robot interaction [11]. Its basic idea is to show specific effects within a system to users, who are then asked to perform gestures that should trigger these effects. The gesture performances are recorded and later analyzed to find gesture candidates that represent the choice of a majority of users. We adopted this approach for Traveller, in which full body gestures performed by the users trigger in-game actions that can vary in their type. For the present purpose we focus on two common action types: a) *navigation*, i.e. changing the position and perspective of the virtual camera and b) *dialogue* with virtual agents.

In the next section, we describe the scenario, including the story script and learning goals of our system, and the architecture used to model the agents' behavior. In Section 3, the development of the user-defined gestures set is depicted, which represents the main part of the interaction within our system. Afterwards, a secondary interaction type for cases in which the main interaction is hard to apply is presented, followed by a conclusion.

2 Scenario

2.1 Background Story and Learning Goals

In Traveller, users take the role of a character that has never been abroad. Throughout the game, they follow in the footsteps of their deceased grandfather, who used to travel the world. Their grandfather has left a letter, to be opened on the character's 18th birthday, in which he states that he has hidden a treasure long ago. To keep it safe, he has left pages of his travel journal in a few countries that he used to visit. To find these pages, the users have to travel to three different countries, and interact with the inhabitants of those countries in so-called critical incidents. The journal pages tell the users where to go to next, and also describe their grandfather's experiences as a beginning traveller.

The story starts at their grandmother's café, where the grandmother gives initial instructions. Afterwards, the users travel to the first country, in which they have to get directions from a group of strangers in a bar (first critical incident), have to find the responsible supervisor in a nearby museum in order to receive entry permission for a

[1] `http://www.xbox.com/KINECT`

park (second critical incident), and support or blame the supervisor when he knocks over a priceless artifact (third critical incident). In the second country, the users have to interact with an elderly man who wants to sit in the users' seat, interact with a train conductor who claims they have the wrong ticket, have a stranger join them for dinner and who might steal their wallet, and help out at a café to earn some money. In the third country, the users have to decide whether they care more about finding the treasure, or helping somebody in need, and interact with people at a party. At the end of the game, the users discover that the hidden treasure was actually in the experiences and adventures that they had during their travels.

The characters in each country have different rules for behavior and interpretation, depending on their synthetic culture. Therefore, if users do not select the appropriate actions, the virtual characters can get upset, and be unwilling to help them on their way. This can sometimes lead to obvious misunderstands, or even outright conflicts, but also to situations in which the users do not realize that they have offended the characters. The cultural configuration of the virtual characters is based on synthetic cultures [5]. These synthetic cultures reflect the extremes of Hofstede's dimensions of culture [4], which were empirically validated across several nations. At the moment, we only use three dimensions to determine the behavior of the characters: *power distance*, *individualism* vs. *collectivism*, and *masculinity* vs. *femininity*.

The aim of Traveller is comprised of an affective goal, which focuses on the users' emotions, and a cognitive goal, which focuses on the users' knowledge and understanding. The *affective goal* focuses on making the users aware that their rules for interpretation of appropriate behavior might be incorrect. For example, in certain cultures, women are not willing to talk to a stranger. As a stranger, it is easy to think they are just rude, but it might just be that it is inappropriate for women to talk to a stranger in public. Self-reflection and articulation of how the other person's behavior makes one feel then allows the users to recognize their emotions as they arise in reaction to such a novel situation. These emotional responses may not always be positive in the first place, but should be accepted and integrated without feelings of prejudice. The *cognitive goal* focuses on making the user understand general differences in cultures. As there are similar actions that occur within each country, the users are able to see the effect of similar actions in different countries. For example, a stranger would be treated differently in a collectivistic culture, than in an individualistic culture. By experiencing similarities and differences between cultures, the users see how various standpoints of cultural groups lead to respective behavior and assumptions. The attempt to see the same action from the point of view of people from another culture in turn prompts perspective taking and provides the foundation for empathic responses.

2.2 Cultural Agent Architecture

The reasoning and behavior of the characters in Traveller is driven by the Social Importance Dynamics (SID) model [10], which was integrated in the FAtiMA agent architecture [1]. The SID model is an adjustable model of cultural influences in social behavior that is based on Kemper's status-power theory [6]. The model augments the standard BDI agent framework [3] with a set of social dynamics that constitute human socio-cultural behavior. More specifically, it models the human notion that others, from a relational perspective, are more or less important and that importance determines how much we are willing to act in their best interests. Conversely, it also determines how much we feel entitled to have others act in our favor. To give an example, if a close friend or a family member needs a place to stay overnight, we would gladly provide a room for him or her to sleep. However, the same is less likely to happen if the person is a stranger. Still, if the stranger simply asked for directions to a hotel, then most of us would comply.

Cultures greatly differ on how much social importance (SI) is attributed to others and how much it is conveyed by certain actions. For instance, in many Western cultures, cheek kissing is a very common greeting between acquaintances of the opposite sex. Oppositely, in China it is considered to be a very intimate act. The SID model enables the representation of such conventions. To illustrate how the model has been applied in Traveller, we consider the second critical incident of the first country that takes place in a museum, and in which the user is trying to find a supervisor of a wild park to ask for entry permission. If the user chooses to approach the supervisor directly, the supervisor's response will depend on his cultural configuration. One dimension of this configuration is *power distance*, which indicates how people treat others with higher status. If the supervisor's culture has a high *power distance*, he will not accept a direct request from the user with lower status. If the *power distance* score is low instead, the supervisor will directly accept the request.

Besides the SID model, the architecture also features the capacity to synthesize emotions in response to events that happen in the virtual environment, following the OCC appraisal theory [12]. This is essential for making the characters seem believable and for the user to establish an empathic relation with them. The architecture has also two different layers to control the behavior of the agents, a Reactive layer and a Deliberative Layer. The first one is responsible for generating quick reactions to events, such as a facial expression triggered by an emotion. The second one endows agents with goal-oriented behavior. These capabilities were already existent in FAtiMA and more details about them are described in [2].

The virtual environment for the scenario was implemented using the cross-platform game engine Unity3D[2]. The ION framework [13] is used to manage the communication between the agent architecture and the 3D world. In addition, characters speak by using the Microsoft Speech API and voices from CereProc[3].

3 Development of User-Defined Gestures

For the development of our user-defined gesture set, we conducted a study that is described in the following section. In the subsequent section, we present the resulting gesture set and how we integrated it into our application.

---

[2] http://unity3d.com

[3] http://www.cereproc.com

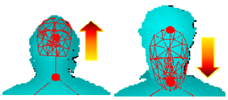
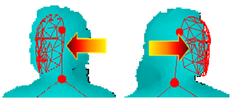
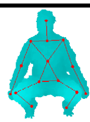
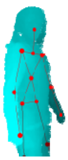
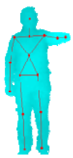
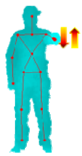
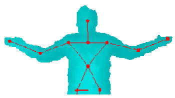
3.1 Gesture Study

We conducted our gesture study for the introduction scenario and the first two critical incidents within the first country. They included the following in-game actions to be triggered via body gestures: *yes, no, sit at bar and wait, approach group, ask for directions, leave the bar, ask about supervisor, ask guard to talk to supervisor, approach supervisor, ask permission.*

Whenever an interaction was requested by the system, we displayed all available actions as text boxes on the screen and the users' task was to invent and perform a gesture for each displayed option, one after the other. Those gesture performances were recorded on video and analyzed later to find gesture candidates for the investigated actions. The gesture candidates were chosen according to an agreement score based on how often users performed the same gesture for one action. In this way, we got one or two gesture candidates per action depending on the level of agreement between the participants. In the case of two gesture candidates we needed to decide which one we would use in the final gesture set. This was done in a way that the single gestures fitted to each other, and there was no problem of ambiguous gestures in parallel.

The final gesture set and its integration is described in the next chapter. Further details about the gesture study, the results of its analysis, and the implementation of the gesture recognition can be found in [7].

Table 1. Implemented gesture candidates and related actions (in brackets)

 head nod *(yes)*	 head shake *(no)*	 sit down *(sit at bar and wait)*		
 step forward *(approach group)*	 turn away *(leave the bar)*	 point to front *(ask guard to talk to supervisor)*	 tip on shoulder *(ask permission)*	 arms out *(ask for directions, ask about supervisor)*

3.2 Gesture Set and Its Integration

For integrating the gestures in our application we use the Kinect for Windows SDK[4] together with our FUBI framework of which an earlier version has already been presented in [8]. FUBI achieves gesture recognition by using an XML-based definition language. To instruct users, we display symbols (single images or animations) that visualize how the gestures for these actions should be performed. The gesture

[4] `http://www.kinectforwindows.org`

candidates we extracted in our study and integrated in the first part of our scenario are depicted by their onscreen symbols in Table 1. Most of them are actually animations, but only one or two important key frames are displayed for reasons of clarity. As soon as a symbol is displayed on screen, the recognition framework automatically checks the corresponding recognizer for the closest user in the depth sensor's field of view. In case the recognition has been successful, it triggers an event related to the symbol in the same way as it is done for default interface buttons in Unity3D. Fig. 1 depicts a scene of the first critical incident displaying four symbols of the new gesture set.

Fig. 1. Gesture symbols in the first critical incident

4 GUI Interaction

During the further development of Traveller's story, it became clear that our scenario would include interactions that sometimes have multiple conversational actions in parallel that would be hard to represent with unambiguous gestures. Therefore, we decided to add a second type of interaction to our application.

In case many conversational actions are necessary at a specific point in time, we group those actions into a dialogue menu as shown in Fig. 2. For entering the dialogue menu, users have to perform the "arms out" gesture as depicted in Table 1. When the dialogue menu is available, the symbol for the "arms out" gesture is shown in parallel to the other currently available actions, which are directly represented by gesture symbols. When the users perform the "arms out" gesture, the dialogue menu opens with the additional available conversational actions and one option to close the menu again. Within the menu, the options are arranged around a circle in the middle of the screen, with each of them occupying an equally sized sector around the middle circle.

For selecting one of the options in the dialogue menu, users first have to stretch out their hand to the front, wait until the menu gets activated, and then perform a swiping gesture in the direction of the option they would like to select. Activation of the menu is visualized by the middle circle changing its color from blue to yellow. In addition, the circle always contains textual instructions for what to do next. As soon as the start of a swiping gesture is recognized, the corresponding arrow gets a little stretched in its pointing direction, and also changes its color to yellow together with the background of the corresponding action text. As soon as the swipe is completed and thus a selection is performed successfully, a sound is played for additional feedback.

In this way, we keep the freedom to develop the story with as many and complex actions as we want, without worrying about how all of them could be represented by unambiguous gestures. However, the interaction modality remains the same for all in-game actions, and the two interaction types are similar enough to provide a fluent user experience.

Fig. 2. GUI Interaction menu

5 Conclusion

In this paper, we presented a novel approach to culture training that is based on role play with virtual characters. By engaging in interactions with characters that simulate different synthetic cultures, users may actively experience the challenges of cultural communication. Particular emphasis was given to the design of natural forms of gesture-based interaction to achieve the required social immersion which we consider as a decisive factor of success for social learning. First informal studies with the system at public events have shown that users are very engaged with the system, eager to explore the scenarios and enjoyed the gesture-based interaction. Future user studies will investigate to what extent interaction with the system may foster cultural awareness.

Acknowledgments. This work was funded by the European Commission within FP7 under grant agreement eCute (FP7-ICT-257666).

References

1. Dias, J., Mascarenhas, S., Paiva, A.: Fatima modular: Towards an agent architecture with a generic appraisal framework. In: Proc. of the Int. Workshop on Standards for Emotion Modeling (2011)
2. Dias, J., Paiva, A.: Feeling and reasoning: A computational model for emotional characters. In: Bento, C., Cardoso, A., Dias, G. (eds.) EPIA 2005. LNCS (LNAI), vol. 3808, pp. 127–140. Springer, Heidelberg (2005)
3. Georgeff, M.P., Ingrand, F.F.: Decision-making in an embedded reasoning system. In: Proc. IJCAI 1989, pp. 972–978 (1989)
4. Hofstede, G., Hofstede, G.J., Minkov, M.: Cultures and organizations: Software of the mind: Intercultural cooperation and its importance for survival, 3rd edn. McGraw-Hill, New York (2010)
5. Hofstede, G.J., Pedersen, P.: Synthetic cultures: Intercultural learning through simulation games. Simulation & Gaming 30(4), 415–440 (1999)
6. Kemper, T.: Status, power and ritual interaction: a relational reading of Durkheim, Goffman, and Collins. Ashgate Publishing Limited, England (2011)
7. Kistler, F., André, E.: User-defined body gestures for an interactive storytelling scenario. In: Kotzé, P., Marsden, G., Lindgaard, G., Wesson, J., Winckler, M. (eds.) INTERACT 2013, Part II. LNCS, vol. 8118, pp. 264–281. Springer, Heidelberg (2013)
8. Kistler, F., Endrass, B., Damian, I., Dang, C., André, E.: Natural interaction with culturally adaptive virtual characters. Journal on Multimodal User Interfaces 6, 39–47 (2012)
9. Kurdyukova, E., Redlin, M., André, E.: Studying user-defined iPad gestures for interaction in multi-display environment. In: Proc. IUI 2012, pp. 1–6 (2012)
10. Mascarenhas, S., Prada, R., Paiva, A., Degens, N., Hofstede, G.J.: Can i ask you a favour? - a relational model of socio-cultural behaviour. In: Proc. AAMAS 2013. Springer, Heidelberg (2013)
11. Obaid, M., Häring, M., Kistler, F., Bühling, R., André, E.: User-defined body gestures for navigational control of a humanoid robot. In: Ge, S.S., Khatib, O., Cabibihan, J.-J., Simmons, R., Williams, M.-A. (eds.) ICSR 2012. LNCS, vol. 7621, pp. 367–377. Springer, Heidelberg (2012)
12. Ortony, A., Clore, G., Collins, A.: The Cognitive Structure of Emotions. Cambridge University Press, UK (1988)
13. Vala, M., Raimundo, G., Sequeira, P., Cuba, P., Prada, R., Martinho, C., Paiva, A.: ION framework – a simulation environment for worlds with virtual agents. In: Ruttkay, Z., Kipp, M., Nijholt, A., Vilhjálmsson, H.H. (eds.) IVA 2009. LNCS, vol. 5773, pp. 418–424. Springer, Heidelberg (2009)
14. Wobbrock, J.O., Morris, M.R., Wilson, A.D.: User-defined gestures for surface computing. In: Proc. CHI 2009, pp. 1083–1092. ACM, New York (2009)

Usability Guidelines for Mobile Devices: Length of Hyperlinks

Eva Garcia-Lopez, Antonio Garcia-Cabot, Luis de-Marcos, and Jose-Ramon Hilera

Computer Science Department, University of Alcalá,
Ctra. Madrid-Barcelona Km 33.6, 28871, Alcalá de Henares, Madrid, Spain
{eva.garcial,a.garciac,luis.demarcos,jose.hilera}@uah.es

Abstract. The use of mobile devices is increasing in recent years, and their characteristics are different from those of computers (keyboard and screen size, interaction method, etc.), so it is important to specifically study the usability of these devices. This paper presents an experiment carried out with users to find the correct length (from the point of view of usability) of hyperlinks in mobile devices. A real mobile device has been used for the experiment, as well as more than 20 users. The experiment compared three different lengths of hyperlinks and it concluded that users prefer hyperlinks as short as possible and they are identified better when the text is not justified.

Keywords: Hyperlinks, usability, guidelines, mobile devices.

1 Introduction

The rise of mobile devices has led to a transformation in the way of creating digital content, not just web pages but any content that these devices can play [1]. Some characteristics of these devices, such as the small screen size or the way of interaction, pose a real challenge when generating this content for mobile devices [2].
Focusing on web content development from the point of view of usability, there are several efforts to establish guidelines and/or recommendations to generate this content in a usable way [3].

Some important aspects of the web, by its nature, are hyperlinks that allow linking some content with another one. From the point of view of traditional computers, there have been many efforts to establish guidelines and recommendations to create these hyperlinks (their size, significance, etc.); but from the point of view of web content for mobile devices, perhaps these guidelines cannot be used or should be adapted by taking into account the limitations of these devices.

Therefore, this paper presents a solution to the problem of which length should hyperlinks have when they are shown in mobile devices, always from a usability perspective.

To solve the problem, this paper presents a set of experiments carried out with users and with different kinds of hyperlinks that allow figuring out the right size for them.

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 705–712, 2013.

Section 2 of this paper shows the background and the related work to this research, Section 3 explains the details of the experiments carried out, Section 4 presents the results and finally Section 5 shows the conclusions and future work.

2 Background and Related Work

Some guidelines and standards establish a concrete length for hyperlinks. In some cases the link length is not explicitly mentioned, but a link should be understandable enough, regardless of its length. Furthermore, some of these guidelines were established for web pages in general, such as [4], [5]. However, other guidelines have already been defined specifically for mobile devices [6], [7].

The standard ISO 9241-151 [4], designed specifically for general purpose web pages, states in guideline 9.4.13 “Link length” that “Textual link names should be long enough to be understood but short enough to avoid wrapping”.

The initiative Usability.gov [5] states in guideline 10:11 “Use Appropriate Text Link Lengths” the following: “Make text links long enough to be understood, but short enough to minimize wrapping”, which somewhat matches the ISO guideline about creating links long enough to be understood but preventing wrapping.

Furthermore, Budiu, R. and Nielsen, J. [6] propose a set of guidelines focused on web and applications design for mobile devices, and in guideline 103 they state “Use links with good information scent (that is, links which clearly indicate where they take the users) on your mobile pages”. The length that links should have is not explicitly shown, but it indicates they must be clear and descriptive.

Finally, W3C [7] has best practices for creating web pages, and guideline 5.2.6 “Link Target Identification” states that links should clearly indicate the target of the link, but does not mention any length that links should have.

Due to the above, this paper presents a set of experiments with real users to help making a recommendation about the importance of the size of hyperlinks in mobile devices. The hypothesis will be taken from the aforementioned guidelines of ISO 9241-151 and Usability.gov, i.e., “Texts of the links should be long enough to be understood and short enough to avoid wrapping” because, although they are not guidelines applied directly to mobile devices, they are concrete and specific.

3 Planning the Experiment

An experiment was designed to test the usability of different link lengths using metrics from the “usability” concept according to ISO 9241-11 [8] (effectiveness, efficiency and satisfaction). For this, three types of web pages (Fig. 1.) were designed, each of which contained a text with some words or sentences marked as links. These links were colored and underlined, as suggested by [9], to maximize the perceived affordance of clickability. The text used was extracted from a chapter of The Theory of Evolution by Charles Darwin.

Links in the first web page were short, i.e., each one of them consisted of a single word. In the second web page links were medium, i.e., they had at least two words but without taking up more than two lines. Finally, links in the third web page were long, occupying several lines.

Fig. 1. Prototype of web pages designed for the experiment: short, medium and long size links, respectively (in Spanish)

Users were asked to count the number of links in each of the three web pages, so that the **effectiveness** could be measured objectively, using the following formula:

$$effectiveness = 1 - \frac{|n - i|}{n} \tag{1}$$

where n is the number of real links and i is the number of links the user counted.

It is noteworthy that the links were disabled, that is, if the user clicked on it, no action would be triggered. This is to prevent the user found out if, for example, a long link was composed of a single link or more than one.

Another parameter to measure was the time that users took to count the number of links. In this way, **efficiency** could be measured, taking into account the average time taken in counting the number of links in each of the cases.

Finally, after the experiment, users completed a **satisfaction** survey in which they were asked about how comfortable was for them the act of reading the text with each type of links, and if the target of the links was easily understood in each web page of the experiment. The survey had also some "free response" questions about the opinion of each web page, so we could infer what users think, in general, about each of the types of links (short, medium and long size).

In order that the text of the web page was not the same in all treatments, three different versions of web pages were designed with the three types of links in each, so that there were nine different web pages for distribution to users. The number of links for each web page was different. Treatments were randomly distributed among users, so that each user performed the experiment with a web page with short links, another one with medium size links and finally one with long size links, but never with the same text.

Additionally, other three web pages were created with the same text and with different links (long size), but with justified text instead of left-aligned. This is because when the experiment was performed with "subject 0" (the first user that participated in the experiment), we realized that the justification (or not) of text could confuse the user when counting the links, especially when they take up more than one line, as it is the case of long size links.

4 Results and Discussion

The experiment was performed with 22 subjects, of which 77.27% were male and 22.73% female. 63.64% of the subjects were between 18 and 24 years; 22.73% were between 25 and 34 years; and 13.64% were older than 35 years. Of these, 45.45% was considered expert in the use of mobile devices, 45.45% was considered in an intermediate level, and 9.09% was considered novice. Regarding the highest level of education completed, 13.64% of the subjects had High School completed and were studying a Degree; 31.82% had Associated Degree; 18.18% had a Degree; 31.82% had a Master Degree and 4.55% had a PhD.

Fig. 2. Mechanism used to record user interactions

To carry out the experiment, a mobile device of make Apple was used (specifically model iPhone 4) and also a webcam that recorded continuously the screen that the user was seeing. The webcam was joined to a metal structure, and the structure was joined to a soft housing specific for this kind of device. The webcam was connected to a laptop via the USB cable, so the videos were recorded directly into the laptop.

Four videos were recorded for each user, one for each factor:

- One with unjustified text and short size links.
- Another with unjustified text and medium size links.
- Another one with unjustified text and long size links.
- Another one with justified text and long size links.

To measure **effectiveness**, as mentioned above, it was measured by the percentage of correct answers of the subjects to count the number of links in each configuration, that is, with short size links, with medium size links, with long size links and unjustified text, and with long size links and justified text. Then the average of successes of all subjects for each configuration and the standard deviation were obtained (Table 1).

Table 1. Results obtained for effectiveness

Type of link	Average (% Success)	Standard Deviation	Median
Short size	99.68	0.0152	1
Medium size	100	0.0000	1
Long size and unjustified text	98.82	0.0312	1
Long size and justified text	97.30	0.0465	1

Later a normality test for data was executed, in order to determine the most appropriate method of analysis (Table 2).

Table 2. Normality test for effectiveness

Type of link	*p*
Short size	<0.005
Medium size	NA*
Long size and unjustified text	<0.005
Long size and justified text	<0.005

*All values are 1 (100% success) for medium-sized links so *p* cannot be calculated.

Data did not pass the normality test (p<0.05) so nonparametric methods should be used. Kruskal-Wallis test returns p=0.009 (<0.05; H(3)=11.46), so the results are statistically significant. That is, the size of the links affects the effectiveness. To distinguish those groups that are statistically better and worse than others, another Kruskal-Wallis test was applied, but this time by pairs of groups (Table 3).

Table 3. Results of Kruskal-Wallis test

Type of link	*p*	H
Short and medium sized links	0.317	1
Short links and long links with unjustified text	0.290	1.12
Short links and long links with justified text	0.022	5.23
Medium links and long links with unjustified text	0.076	3.14
Medium links and long links with justified text	0.004	8.07
Long links with unjustified text and long links with justified text	0.195	1.68

In view of the results, short and medium sized links are statistically more effective (i.e., users make fewer mistakes) than long links with justified text. Moreover, one can say, with 90% confidence, that medium sized links are statistically more effective than long links with unjustified text.

Furthermore, for **efficiency**, we analyzed the time each subject took to count the number of links with each configuration, and the average, standard deviation and median were obtained for each (Table 4).

Table 4. Results obtained for efficiency

Type of link	Average (seconds)	Standard Deviation	Median
Short links	14.091	6.286	12
Medium links	15.773	6.164	13
Long links and unjustified text	26.455	12.054	21
Long links and justified text	23.273	7.857	22

First of all, a normality test was applied to data, in order to determine the most adequate analysis method (Table 5).

Table 5. Normality test for efficiency

Type of link	p
Short links	<0.005
Medium links	0.009
Long links and unjustified text	<0.005
Long links and justified text	0.222

Data did not pass the normality test ($p<0.05$), so nonparametric methods should be used. Kruskal-Wallis test returns $p=0.000$ (<0.05, $H(3)=29.53$), from which it follows that the results are statistically significant, i.e., short and medium sized links are statistically more efficient (faster to identify) than long size links. It is not possible to say that there is statistical significance between short and medium sized links; nor between long links with unjustified text or with justified text, so it cannot be said that some are more efficient than others.

Finally, the surveys that users answered after the experiment were analyzed to know the **satisfaction** of the users. Answers to three questions were assessed, which were the following for each configuration:

- Q1: Is it comfortable to read the text?
- Q2: Is the target of the link easy to understand?
- Q3: What do you think, in general, you will get if you click on the links?

The answers to the first two questions were measured quantitatively on a Likert scale from 0 to 5. The results for each configuration (average of answers of all users and standard deviation) are shown in Table 6.

Table 6. Survey results

Type of link	Q1 (Average, from 0 to 5)	Q1 (Standard deviation)	Q2 (Average, from 0 to 5)	Q2 (Standard deviation)
Short links	4.36	0.726	3.27	1.579
Medium links	3.27	1.077	3.05	1.090
Long links and unjustified text	1.41	1.007	2.36	1.733
Long links and justified text	1.55	1.299	2.59	1.436

As for the third question, answers of the subjects were "free", so it is difficult to quantify. However, much of the users (90.91%) think that short links will lead to a web page where the definition of the word is explained, while long and medium sized links usually make to think (by 77.27% of cases) that will lead to a web page that explains in more detail the link sentence.

5 Conclusions and Future Work

In view of the results above, the following conclusions are obtained:

- Looking at the comfort for reading a text of a web page from a mobile device (Q1), there is no significant different between justifying or do not the text, that is, some users think that reading the text is more comfortable when it is justified but some others think it is easier when the text is not justified, since there is no a big difference between their averages. However, efficiency is influenced negatively by the text justification, i.e., when the text is justified, users seem unable to differentiate where a link starts and where it ends as well as when the text is not justified. Therefore, one could argue that **it is preferable do not justify the text** so that the user can clearly distinguish where the links start and end, although this affirmation is based on the average of data and it must be taken into account that it is not statistically significant.
- Users prefer clearly short links (according to Q1 average), because they think that is the most comfortable way to read a text containing links. Medium links are not as uncomfortable as long links, and observing an increasing score in reading comfort from long links to short links, it follows that users seem to **prefer links as short as possible**. This conclusion is further supported because data suggest that users often understand better the target of a link (according to Q2 average) the shorter (there is an increasing score about understanding the target of the link, from the longest to the shortest), and also because effectiveness is higher the shorter the link (users seem to spend less time when identifying the links). As a remark, note that the effectiveness is slightly better with medium links than with short links, but this does not really matter, since only one user was confused when he counted short links. Therefore, the conclusion “the shorter the links the better” is still valid despite this fact.

- Comparing the above conclusions with the **initial hypothesis** ("Texts of the links should be long enough to be understood and short enough to avoid wrapping"), one could say that in the case of mobile devices this **is not true** because users seem to prefer links as short as possible, and also the data suggest that they are better and faster identified, both where they start and end, and their target. Therefore, a possible recommendation about the link length in web pages for mobile devices would be: "Texts of the links should be as short as possible and, whenever possible, the text in which they are contained should not be justified".

For future work, the experiment could be repeated with different mobile devices with diverse characteristics (techniques of interaction, screen size, screen orientation), in order to check if conclusions can be extrapolated to other devices aside from smartphones. Also, in addition to the length, the semantics of links should be considered to evaluate the understanding of the links by users.

Acknowledgement. This research has been co-funded by the Master in Software Engineering for the Web of the University of Alcalá, the projects IPT-2011-0795-430000 and UAH2011/6XP-0043090XT099, and the TIFyC research group.

References

1. Gómez, S., Fabregat, R.: Context-Aware Content Adaptation in mLearning. In: Montebello, M., Camilleri, V., Dingli, A. (eds.) 9th World Conference on Mobile and Contextual Learning, MLEARN 2010, pp. 76–83 (2010)
2. Zhang, D., Adipat, B.: Challenges, Methodologies, and Issues in the Usability Testing of Mobile Applications. International Journal of Human-Computer Interaction 18(3), 293–308 (2005)
3. Nielsen, J., Budiu, R.: Mobile usability. New Riders, Berkeley (2012)
4. ISO. ISO 9241-151:2008. Ergonomics of human-system interaction – Part 151: Guidance on World Wide Web user interfaces (2008)
5. U.S. Department of Health and Human Services'. Research-Based Web Design and Usability Guidelines (2002), `http://usability.gov/guidelines` (last accessed March 22, 2013)
6. Budiu, R., Nielsen, J.: Usability of Mobile Websites and Applications, Design Guidelines for Improving the User Experience of Mobile Sites and Apps. Report, Nielsen Norman Group (2011)
7. W3C. W3C Mobile Web Best Practices 1.0 (2008), `http://www.w3.org/TR/mobile-bp/` (last accessed March 22, 2013)
8. ISO. ISO 9241-11:1998. Ergonomic requirements for office work with visual dis-play terminals (VDTs) – Part 11: Guidance on Usability (1998)
9. Nielsen, J.: Jakob's Nielsen Alertbox: Guidelines for Visualizing Links (2004), `http://www.nngroup.com/articles/guidelines-for-visualizing-links/` (last accessed March 24, 2013)

User-Centered Design between Cultures: Designing for and with Immigrants

Jan Bobeth[1], Stephanie Schreitter[1], Susanne Schmehl[1],
Stephanie Deutsch[1], and Manfred Tscheligi[1,2]

[1] CURE – Center for Usability Research and Engineering,
Modecenterstr. 17/2, 1110 Vienna, Austria
[2] ICT&S Center, University of Salzburg, Sigmund-Haffner-Gasse 18,
5020 Salzburg, Austria
{bobeth,schreitter,schmehl,deutsch,tscheligi}@cure.at

Abstract. Immigrants represent a substantial part of European society. After emigration, they can suffer from fundamental changes in their socio-economic environment. Therefore, supportive ICT services (e.g. for language learning or job search) have high potential to ease inclusion, especially for newly arrived immigrants with low education. Within an international research project we involve Turkish and Arabic immigrants in a user-centered design (UCD) process with the goal to develop supportive ICT services for smartphones. In this paper, we present our methodological experiences and discuss benefits and drawbacks of methods. Based thereupon, we formulate concrete implications for successful UCD with immigrants, e.g. collaborating with nongovernmental organizations (NGOs) or benefiting from reflections of long term-immigrants.

Keywords: User-Centered Design, Immigrants, Diversity, Method.

1 Introduction

Migration is and was always common at all parts of the world. It´s occurrence is usually influenced by a combination of economic, political and social factors. Supporting the social integration of immigrants in Europe and providing them access to labor markets is essential for the prospective welfare of natives and immigrants. For successful social integration immigrants adapt to basic values and participate as an integral part in the host society, while at the same time maintain their original culture [2]. Within an international research project we aim to leverage the potential of mobile ICT services provided via smartphones for supporting social integration by offering ad-hoc support (e.g. navigation or real-time translation) as well as long-term training (e.g. mobile language learning). These services are dedicated to overcome the main barriers a newly arrived immigrant has to deal with, when trying to gain ground in an unfamiliar country.

Immigrants moving to countries within the European Union come from all over the world [8]. Thus, they form a very heterogeneous group and can hardly be treated as

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 713–720, 2013.

one sole target group in terms of service and user interface needs. Their needs differ widely since great variety of user characteristics occurs depending on maternal language, cultural background, education, motivation, duration of stay, religion, and profession. For this reason, it is not expedient to approach all immigrants in Europe at once when designing services for social integration. It is necessary to focus on a narrower definition of the target group. We mainly address Turkish-speaking Turks and Arabic-speaking immigrants from North Africa as they form two large immigrants groups within the European Union [8]. According to the cultural model of Hofstede [11], the cultural background of both groups differs noticeable from Western cultures (mainly in the dimensions Power-Distance and Collectivism/Individualism). Additionally, we focused on immigrants with low education (up to eight years) and rather short length of stay (up to five years) as these immigrants are potentially exposed to more barriers in everyday live.

Research has shown that cultural differences matter in interface design and affect the data-gathering in a user-centered design process [5]. However, for the special case of immigrants it remains unclear which impact cultural differences between home and host country might have on the outcome of user-centered design processes as well as on interface preferences. By putting immigrants at the center of our research we want to assure that services are needed and useable for the respective target group. Following a user-centered design process (UCD) [20], a crucial attempt is to understand the users, their problems as well as their needs and accordant contexts of use. Therefore, the adoption of a hermeneutic approach with qualitative methods seems to be most fruitful for the elicitation of cultural-specific user requirements (e.g. observing and interviewing the concrete target group) [1].

In this paper, we present our methodological experiences gathered within requirements analysis (creating a barriers list, conducting semi-structured interviews) as well as service and interface design (conducting focus groups and participatory design workshops) as parts of the UCD process for and with a selected immigrant user group. We discuss benefits and drawbacks of methods with the goal to formulate concrete implications for successful UCD for and with immigrants.

2 Related Work

While sociologists have been investigating social integration and cultural differences of immigrants for a long time (e.g. [9, 13]), researchers in the domain of human-computer interaction focused mainly on comparing influences of different cultures (e.g. [5]) instead of targeting immigrants living "between" two cultures. In this section, we present related work from both perspectives. For the elicitation of cultural-specific user requirements Aykin et al. [1] recommended qualitative methods. However, when applying qualitative methods in the work with immigrants several issues arise, such as vulnerability and mistrust towards researchers. The immigration status is very relevant regarding vulnerability and the status may change. Mistrust towards researchers and their work is a general issue for immigrants. Building up trust can require more than application of anonymity, confidentiality, and the use of ethical principles (e.g. by working voluntarily within a nongovernmental organization) [12]. For these reasons, it is necessary to maintain a flexible research approach [13]. A researcher is obliged to provide any information about the study, its purpose, and data handling to participants to give them the possibility and the power to decide whether to participate or not [14]. Bloch [3] found that asylum seekers were less willing to participate in research than refugees, due to anxieties about repercussions if their responses are given to the local authority. To better understand and reflect the results it is vital to know as much as possible about the study participants (e.g. language and literacy skills, cultural norms, etc.) [3].

Regarding concrete methods, self-completion questionnaires have the advantage of being relatively cheap to administer as well as more suitable than face-to-face interviews when sensitive questions are asked or the research itself is sensitive [6]. Face-to-face interviews might be viable to avoid a lacking of answers due to literacy skills [6]. Interviewers sharing ethnic background and mother tongue with the interviewed immigrants might be more successful with sensitive questions [7]. Talking in mother tongue can also avoid discomfort for the interviewees [4]. Frindte et al. [9] reported about a multi-generation case study in which interviews, surveys, and discussions were conducted in German, Turkish, or Arabic. Using bilingual Turkish-German and Arabic-German interviewers proved to be very effective and reduced the mistrust of the participants concerning the research.

3 Method Framework for User-Centered Design

Based on the experiences from related work, we developed a method framework for involving immigrants. However, a big challenge for UCD with immigrants is to find and recruit real users matching the target group definition [1]. Another problem especially in the first phase of the requirement analysis is the establishment of trust between immigrants and researchers [12]. For this reason we collaborated with three nongovernmental organizations (NGOs) in Graz (Austria), London (UK) and Madrid (Spain) that are in frequent contact with immigrants and were responsible for the recruiting of all study participants. Being introduced to immigrants by co-workers of the NGOs is already a positive sign for certain trustfulness of for immigrants' unknown researchers. All study participants gave informed consent. In the following we describe how we involved end users and NGOs in the UCD process with the help of qualitative methods (see [1]). Applied methods are explained in chronological order whereas content related results are not reported since this is not scope of this paper.

3.1 Barriers List

As contact point for immigrants NGOs have extensive experiences with immigrants' everyday lives, accumulated over years. To learn from these experiences, in the first step of our UCD process co-workers of the NGOs created a list of common barriers for immigrants and related them to age, sex, level of education and length of stay. Lists of the three NGOs were merged and barriers prioritized by frequency of occurrence as basis for setting up the semi-structured interviews conducted subsequently.

3.2 Semi-structured Interviews

By conducting semi-structured interviews, we aimed at gaining deeper insights about barriers and problems in everyday life. Apart from demographic questions, the interview consisted of three parts: (i) questions about social background and qualifications of the interviewees, (ii) motivation for emigration of home country and reasons influencing this decision, (iii) barriers in their personal lives, where they look for support, and whether they could think of mobile services for smartphones supporting them. Bilingual co-workers interviewed the immigrants in their mother tongue (compare [4] or [9]) in facilities of the NGOs, which were familiar to the participants. The interview protocol and the gathered answers have been translated from and to English. In total, 17 persons have been interviewed in Graz, London and Madrid (10 Arabic and 7 Turkish immigrants). Interviewers were allowed to adapt the order of questions and to dig deeper in some topics based on the tenor of the conversation. One part of the interview was the creation of a relationship map (modified of [15]) indicating communication behavior of participants to better understand the media usage for talking to important persons in home and host country (see Fig. 1a and b).

3.3 Focus Groups

Based on the identified barriers, the interview data and previous ideas of involved researchers various service scenarios have been developed. In order to get feedback on the scenarios and to create additional service ideas, we decided to conduct focus groups as they can reveal surprising insights when conversations take their own dynamic way [19]. In contrast to the interviews, we decided to include only immigrants who already lived more than three years in the host country and who know the host language well enough for actively taking part in discussion. We also invited a translator to each focus group to provide language support when necessary. The goal of the focus groups was to discuss service ideas by benefiting from immigrants' experience and retro perspective reflection on their problems. We discussed three service scenarios in four focus groups with overall 30 immigrants in the facilities of the NGOs (three with Arabic immigrants in Graz, London, Madrid and one with Turks in Graz). With respect to the research question, Skop [19] suggested keeping focus groups homogeneous along certain features, such as sex, age, social class, language, etc. Thus, we did not invite Arabic and Turkish immigrants for the same focus group and conducted two of four focus groups only with female immigrants and researchers.

After a short introduction round of researchers and immigrants, we described orally three common barriers out of the barriers list (e.g. needing immediate help in specific situations like filling out forms). After asking and discussing about similar experiences of their own life, we presented three illustrated service scenarios as solutions for the discussed problems by reading the according story aloud (e.g. a service for finding nearby volunteers). Following the presentations, participants estimated whether such a service might have been helpful and what might be improved (e.g. to protect privacy). The focus groups concluded with a general discussion about mobile ICT services as means of support for immigrants. Fig. 1c shows the Arabic focus group in Madrid.

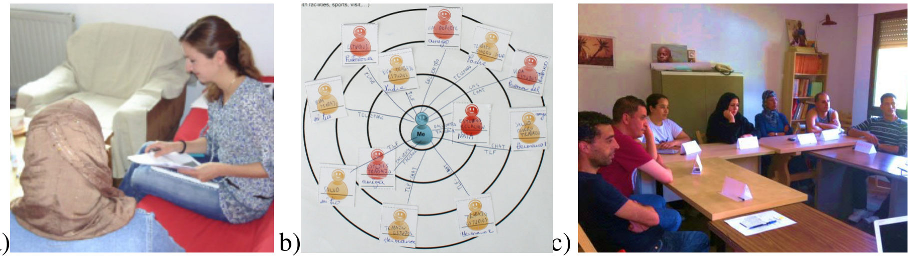


Fig. 1. a) Interview situation with Turkish immigrant, b) Example of a relationship map created during an interview, c) Focus group with Arabic immigrants in Madrid.

3.4 Participatory Design Workshops

To cope with cultural differences in user interface design it is important to actively involve users because most of the existing design guidelines try to raise awareness for the issue but do not offer ready-made solutions [5], especially not for immigrants. For this reason, we conducted two participatory design workshops for co-designing [16] with immigrants and two Austrian designers in Graz and London. The goal was to identify and discuss potential differences in design and solution approaches with respect to the immigration background. We invited Turkish and Arabic immigrants who were sufficiently able to speak the language of the host country and lived there for at least three years. In Graz four Arabic speaking women and four Turkish women took part, while in London 8 participants from Arabic-speaking countries joined.

In order to loosen the atmosphere and to stimulate creativity we started both sessions with an introductory game called Trading Cards [10] replacing the usual introduction round. The tasks were (i) to create a personal trading card including name, self-portrait, nickname and special hobbies within 10 minutes, (ii) to present the trading card of another participant to the audience. Afterwards, two illustrated scenarios with additional written descriptions in the host language were distributed. Groups of two were formed to discuss and work on design ideas for the smartphone application “of their dreams” (compare [17]) for one of the two scenarios that each group could choose freely. Therefore, they received smartphone stencils as well as stickers, pencils, markers, etc. After 30 minutes each group presented the created design ideas in front of the whole group for earning feedback. Subsequently, each group of participants redesigned their ideas in additional 30 minutes and presented their final designs.

4 Discussion and Implications

We present implications of our studies by discussing our experiences with the methods coping with the different cultural backgrounds of researchers and target group.

Collaborating with NGOs. Working in tight connection with the NGOs was of crucial importance. They already had great implicit and explicit knowledge about barriers for immigrants and problems in their daily life. Another crucial aspect is the trustworthy relation between NGOs’ associates and the target group. Furthermore, communication in the immigrants' mother tongue - which is possible at most NGOs - allows avoiding misunderstandings and is an important factor for creating trust.

Supporting Recruiting Process. Recruiting participants according to several criteria is a challenging and time-consuming activity, especially for NGOs working with vulnerable groups like immigrants. Trust and motivation may differ widely within the target group. Some immigrants refuse to interact with foreigners in general, while others might not accept that they cannot take part in a study due to formal criteria (compare [13]). For this reason, we kept the number of exclusive characteristics low (see Introduction) although from a sociological point of view this might not be the most accurate way to e.g. provide statistically meaningful statements. Still, some recruited participants did not match all criteria (e.g. level of education).

Benefiting from NGOs Experiences. Governmental institutions offer mainly statistical data about immigration problems. In contrast, the barriers list created by the NGOs provided a first authentic overview about everyday problems of immigrants. Such a list can be created with few resources and allows a first prioritization of problems to be addressed. A drawback of this approach lies in the choice of the involved NGOs as their co-workers might have a biased view on the problems of their clients. For better understanding concrete problems of immigrants, we originally intended to conduct cultural probing which has been applied successfully by other researchers (e.g. [18]), though with less vulnerable participants. In our case, the involved NGOs recommended to rely on methods that provide immediate answers as vulnerable groups such as short-term immigrants might have too many other problems than documenting their days. Following the recommendations of the NGOs we conducted semi-structured interviews with immigrants in their mother tongue instead.

Interviewing in Mother Tongue and More Implicit Information. Interviews allowed us to gain deeper insights into living circumstances and about potential service needs. Interviewing in the mother tongue of the immigrants created a pleasant atmosphere and allowed easy communication. Probably we would have missed interesting anecdotes when insisting on the host country language and creating discomfort for the interviewee [4]. A disadvantage of interviewing in the participants' mother tongue is the possibility of translation errors (see [13]), which might be increased by the variety of Arabic dialects. As the main goal of the interviews was to get an impression on living circumstances, vague elements in few translations did not distort the data momentously. Nevertheless, to gather more implicit information about daily routines as input for service ideas methods like cultural probing would have been beneficial.

Benefiting from Reflections of Long-Term Immigrants. For a deeper analysis and discussion of service needs, we conducted focus groups with long-term immigrants. They had experienced most of the targeted problems themselves. Thus, they were better able to suggest solutions that could support short-term immigrants than the direct target group itself. Conducting focus groups in the host language and having a translator for language support proved to be effective. Researchers could actively take part in discussions and when participants had problems with formulations the translator could support them. Nonverbal behavior could be observed directly by the researchers (e.g. emotions towards certain barriers or services).

Taking Gender Issues into Account. Concerning gender composition, participants of the two female focus groups pointed out that it was important for them not having to talk to men. Thus, we recommend considering the sex of participants in the selection process for focus groups with immigrants of Turkey and Arabic-speaking countries.

Fostering Openness and Creativity with Playful Methods. In the participatory design workshops we experienced the "icebreaker" game Trading Cards working well. Participants got to know each other and the first barrier to talk to the group fell. The game would have been beneficial for the focus groups as well. Basing the discussion on illustrated scenarios facilitated communication between participants and researchers as they concretized the abstract service ideas. The workshops provided little impact on design ideas, but the interaction with immigrants in face-to-face situations was still valuable for the participating designers to understand how immigrants approach design. It might have been beneficial to set a clearer focus on the conceptual design instead of letting participants try to create concrete designs.

Involving Researchers with Immigration Background. A special challenge of working with immigrants in their mother tongue is the translation of research protocols and gathered data. Co-workers of NGOs are not trained in conducting interviews and translating the answers. Thus, they might miss relevant details, and valuable information might get lost. The best solution could be to collaborate not only with NGOs but also with researchers and designers with immigration background or who are based in the countries of origin. This would be especially interesting to better deal with cultural issues in interface design.

5 Conclusion and Next Steps

To involve vulnerable groups like newly arrived immigrants in UCD is a challenging task. With the help of NGOs as trustworthy partners for immigrants and researchers, we successfully applied a number of methods for analyzing service and interface needs. A list of common barriers created by the NGOs and semi-structured interviews with immigrants supported the researchers to get an overall picture of common problems and to get an impression about living situations. Focus groups and participatory design workshops helped to shape and prioritize service ideas and to better understand design issues. Aim of this paper was to present our implications for other ICT projects involving immigrants in a UCD process. The next steps in the UCD process are the iterative interface design including several usability studies with immigrants and field trials to evaluate the final services. Apart from design challenges like designing Arabic user interfaces also methodological challenges such as investigating influences on usability testing and long-term involvement of immigrants need to be explored.

Acknowledgements. This work has been partly funded by the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n°288587 MASELTOV.

References

1. Aykin, N., Honold Quaet-Faslem, P., Milewski, A.E.: Cultural Ergnomics. In: Salvendy, G. (ed.) Handbook of Human Factors and Ergonomics, pp. 1418–1458. John Wiley, Hoboken (2006)
2. Berry, W.B.: Immigration, acculturation, and adaptation. Applied Psychology: An International Review 46(1), 5–68 (1997)
3. Bloch, A.: Methodological challenges for national and multi-sited comparative survey research. Journal of Refugee Studies 20(2), 230–247 (2007)
4. Blomberg, J., Burrell, M.: An Ethnographic Approach to Design. In: Sears, A., Jacko, J.A. (eds.) The Human-Computer Interaction Handbook, pp. 965–988. L. Erlbaum Assoc., Hillsdale (2009)
5. Callahan, E.: Interface design and culture. Annual Review of Information Science and Technology 39(1), 255–310 (2006)
6. De Leeuw, E.D.: To mix or not to mix data collection modes in surveys. Journal of Official Statistics 21(2), 233–255 (2005)
7. Elam, G., Fenton, K.A.: Researching sensitive issues in ethnicity: lessons from sexual health. Ethnicity and Health 8(1), 15–27 (2003)
8. Eurostat Migration & Migrant Population Statistics, `http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Migration_and_migrant_population_statistics` (last access: March 23, 2013)
9. Frindte, W., Boehnke, K., Kreikenbom, H., Wagner, W.: Lebenswelten junger Muslime in Deutschland. Ein sozial- und medienwissenschaftliches System zur Analyse, Bewertung und Prävention islamistischer Radikalisierungsprozesse junger Menschen in Deutschland (2012)
10. Gray, G., Brown, S., Macanufo, J.: Trading Cards. In: Gamestorming. A Playbook for Innovators, Rulebreakers, and Changemakers, pp. 130–131. O'Reilly Media (2010)
11. Hofstede, G.: Culture's consequences: Comparing values, behaviours, institutions, and organizations across nations, 2nd edn. Sage, Beverly Hills (2001)
12. Hynes, P.: The issue of "trust" or "mistrust" in research with refugees. New Issues in Refugee Research, Working Paper 98. UNHCR Evaluation and Policy Analysis, Geneva (2003)
13. Jacobsen, K., Landau, L.: The dual imperative in refugee research: some methodological & ethical considerations in social science research on forced migration. Disasters 27(3), 95–116 (2003)
14. Lammers, E.: Refugees, asylum seekers and anthropologists: the taboo on giving. Global Migration Perspectives 29 (2005)
15. Moser, C., Fuchsberger, V., Tscheligi, M.: Using Probes to create Child Personas for Games. In: Proc. of ACE 2011. ACM (2011)
16. Muller, M.J.: Participatory design: the third space in HCI. In: Sears, A., Jacko, J.A. (eds.) The Human-Computer Interaction Handbook, pp. 1061–1081. L. Erlbaum Assoc., Hillsdale (2009)
17. Sanders, E.B.N.: Generative Tools for CoDesigning. In: Scrivener, S.A.R., Ball, L.J., Woodcock, A. (eds.) Collaborative Design, pp. 3–12. Springer (2000)
18. Schmehl, S., Deutsch, S., Schrammel, J., Paletta, L., Tscheligi, M.: Directed cultural probes: Detecting barriers in the usage of public transportation. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) INTERACT 2011, Part I. LNCS, vol. 6946, pp. 404–411. Springer, Heidelberg (2011)
19. Skop, E.: The methodological potential of focus groups in population geography. Population, Space and Place 12, 113–124 (2006)
20. Vredenburg, K., Isensee, S., Righi, C.: User-Centered Design: An Integrated Approach. Prentice Hall (2001)

Using Graphical Representations to Support the Calculation of Infusion Parameters

Sandy J.J. Gould, Anna L. Cox, and Duncan P. Brumby

UCL Interaction Centre, University College London,
Gower Street, London WC1E 6BT
{s.gould,brumby}@cs.ucl.ac.uk, anna.cox@ucl.ac.uk

Abstract. A variety of medical procedures require arithmetic calculations to be performed. These calculations can be complex and induce errors that can have serious consequences on the ward. In this paper, we consider whether a graphical representation might make these calculations easier. The results of a laboratory experiment are reported in which participants were asked to solve a number of infusion parameter problems that were represented either graphically or textually. Results show that participants were faster but no more accurate in solving graphical problems than they were textual problems. We discuss the need for situated work to be conducted that builds on these initial findings to determine whether the advantages of graphical representations transfer to actual workplace settings.

Keywords: Graphical reasoning, infusion pumps, re-representation, calculation.

1 Introduction

Many tasks on hospital wards require nurses to perform calculations involving rates, volumes and times (e.g., setting up syringe drivers and infusion pumps). Unfortunately, there is a large body of evidence showing that nurses and trainee-nurses have poor arithmetic skills [1–3]; as a result, calculation errors are occasionally made [4]. This has implications for the quality of care patients receive.

A number of approaches have been taken to improving calculation ability. These have ranged from traditional pedagogic efforts [5] – more teaching – to the development of interactive learning environments [6]. Although educational programmes do much to address arithmetical shortcomings, they provide little insight into whether the traditional textual layout of information used during calculations is best suited to the task that nurses face.

An alternative approach that we explore here is to radically change the presentation of the calculation information from textual to graphical. It has been argued that graphs might make the calculation of infusion parameters easier, thus reducing the likelihood of error [7]. However, there has been little work to date that has evaluated whether people are better at solving these kinds of problems when presented with graphical representations than they are with textual representations.

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 721–728, 2013.

The aim of this paper is to determine whether a graphical representation makes it easier for people to make calculations about infusion parameters. We report the results of a lab-based experiment in which participants were given a set of calculation problems, represented either graphically or textually, along with some questions to answer. We wanted to know whether the graphical representations would allow participants to achieve the same or better levels of accuracy in less time than was possible with textual representations.

1.1 Related Work

Performing calculations is a major component of many nurses' routines. For example, infusion pumps, which are used to deliver medication intravenously to patients, need to be programmed with parameter values, usually infusion duration, volume and rate. Sometimes all of these values are given on a prescription or medication bag so they can be easily programmed into the device. However, sometimes there are missing values that must be calculated before they can be programmed [8]. Even when all of the values are provided, infusions can occasionally be disrupted midway through (e.g. because of a battery failure [8]). Such failures require new values to be calculated and then programmed into the pump so that the infusion can be resumed. Reports show that many nurses find these calculations difficult and more worryingly often overestimate their ability to perform them [9]. Hence there is an opportunity to see whether modifications to environmental artefacts might facilitate accurate calculations.

Previous work [10] has experimented with nomograms – one dimensional slide rule-like representations – to make such calculations easier. Empirical results indicate that these representations can significantly reduce the number of errors made in calculation tasks. Although nomograms ease calculations, they do not provide a representation of the complete time course of the infusion, or its current state, which may cause difficulties when restarting disrupted infusions (e.g., because of a battery failure). It has been proposed [7] that representing infusions on two dimensional charts might make it easier for nurses to reason about the setup of infusions, but this proposal has not been tested empirically.

The advantages conferred by re-representing information have been of interest to the areas of cognitive science and human-computer interaction for some time. Larkin and Simon [11] describe the properties of informational and computational equivalence of representations: representations are informationally equivalent if they are constructed from the same underlying data; they are computationally equivalent if inferences can be drawn with equal speed and ease. Zhang and Norman [12] extend these definitions from a distributed cognition perspective, finding that multiple representations of the same task can afford different degrees of external representation, which in turn affects problem-solving efficacy.

External representations are not necessarily diagrammatic and there is a subset of research concerned specifically with reasoning about representations that use Cartesian co-ordinates (i.e. a chart or graph). Peebles and Cheng [14] provide an empirical validation of the equivalencies described previously by Larkin and Simon [11] in the context of chart-based reasoning. In their study, they compared informationally equivalent graphs and showed that constructing graphs according to particular psychovisual properties can result in reasoning efficiencies that can mitigate participant unfamiliarity with a particular representation. However, this does not seem to apply universally: in some cases where participants were required to reason about equivalent representations, familiarity both with the topic and with graphs is a significant factor in interpretation performance [15]. To reduce the effect of familiarity on reasoning performance, this study uses simple two-point graphs, and participants are not required to perform interpretation; instead, they only need to retrieve calculation parameters from the graph. Despite the extensive work in the area of graphical representations, we are unaware of any comparative experimental studies of the performance characteristics of graphical and textual representations of the same information.

We describe an experiment that investigates whether people are able to perform better in resolving arithmetical problems similar to those that nurses perform for infusions when given graphical representations compared to traditional textual representations. The representations used are informationally equivalent: each problem is represented twice, once in graphical form and once in textual form. However, we hypothesise that the representations are not computationally equivalent: using them requires different degrees of effort. Building on the suggestion that graphical representations might make infusion problems easier to deal with [7], we predict that graphical representations will be less computationally intensive (i.e. easier to use) than textual representations. If graphical representations are indeed less computationally intensive, participants should be able to complete the problems more quickly and more accurately than they can with textual representations.

2 Method

2.1 Participants

Twenty-one participants (11 female) with a mean age of 24 years (*SD*=4 years) took part in the study. Participants were recruited from the psychology subject pool at University College London and were paid £10 for approximately one and a half hours of their time.

2.2 Materials and Design

The experiment was made up of 22 infusion problems. These problems specified the total volume to be infused (i.e. the starting volume), the total duration of the infusion, the time elapsed so far and the total volume remaining. Each of these problems was presented twice during the course of the experiment, once textually and once graphically (see Fig. 1). Participants were not informed that the same problems could appear more than once in different forms.

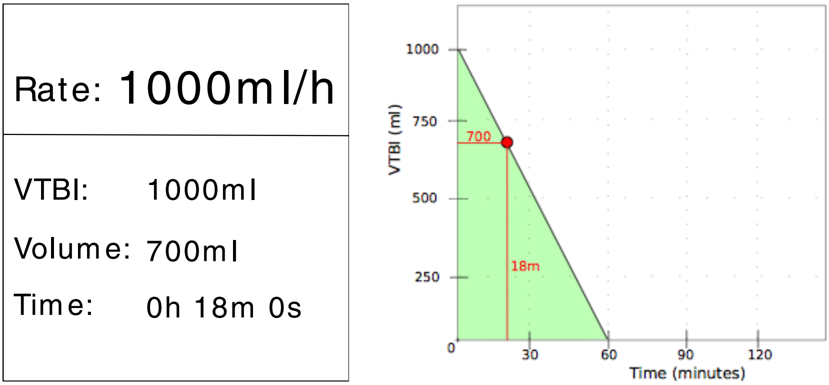


Fig. 1. Examples of the textual (left) and graphical representations used in the experiment

The textual representations (shown on the left of Fig. 1) give the rate of infusion, the total volume (VTBI), the volume remaining and the time elapsed. This representation is typical of the interface of many infusion pumps that are currently manufactured. The graphical representations use the same information, only plotted on a chart. Although the information differs in representation, the total information content is identical. All graphical representations were displayed on the same scale from 0-1000ml/h and 0-120 minutes. In order to encourage calculation, the graphs were not drawn accurately enough to compute answers directly from the graphs by reading values from the axes.

Underneath the representation of the problem, participants were given either one or three questions on the problem, depending on the phase of the experiment. These problems required participants to retrieve information from the representations and then perform a calculation. For example, for the problem shown in Fig. 1, participants might be asked “How many minutes remain in this infusion?” To come to the correct answer – 42 – participants would need to subtract the time elapsed from the total infusion duration. More complex questions might ask, “If the rate of infusion was halved at this moment, how much longer would the infusion take?” Participants answered the questions associated with the problem and then clicked a button to proceed to the next problem. Participants were not given any feedback on their answers.

In addition to the calculation task, participants were also required to complete a cognitively demanding secondary interrupting task. This was a modification of the n-back task [16]. Participants were shown a two-by-two grid. Every three seconds, one of the cells was filled in either blue or red. This was a two-back task so participants had to decide whether the current arrangement matched the arrangement of colour and position they were shown two arrangements prior. The interruptions lasted for 30 seconds after which participants were returned to the primary task.

Finally, an eight item post-experiment questionnaire was also devised. This asked participants about the extent to which they found different aspects of the experiment difficult. Participants indicated their agreement with a statement on a five-point Likert scale.

This study used a counterbalanced within-subjects design with a single independent variable, type of representation, which had two levels, graphical and textual. The dependent variables were answer accuracy; time to complete each problem; and time to answer the first question in a problem after it was presented on the screen.

2.3 Procedure

After reading an information sheet, watching an introductory video and giving consent, participants started working through the training phase. Participants were presented with eight training problems, four each for the graphical and textual representations. Each problem came with three questions, the majority of which required participants to find the appropriate information in the representation and transcribe it into a text field. For example, for the problem in Fig. 1, participants might be asked what volume remained to be infused or how much time was left. Participants completed all of the problems with the assistance of the experimenter. If participants demonstrated satisfactory competence in answering these questions they continued to the experimental trials, which consisted of two phases.

The first experimental phase presented participants with a single question per problem. Participants were given the same problem twice, once with the graphical representation and once with the textual representation. There were a total of nine problems in this first phase, giving a total of 18 trials, two for each problem. The order of presentation was counter-balanced.

The second experimental phase used a different set of nine problems. In order to accommodate interruptions, problems in this phase comprised three questions. Interruptions occurred after the participant had started working on the problem. More specifically, interruptions could occur in one of two places: either before the second question or before the third question. There were a total of 18 trials in this phase, nine of each type of representation. Participants alternated between the two types of representation (i.e. a graphical problem followed by a textual problem) and this order was counterbalanced across participants. Participants were given access to a pen, paper and calculator for the duration of the study and told that they were free to do as they wished with them.

3 Results

Four participants were unable to complete all questions in the time available, so their data were excluded from the analysis. Participants answered a total of 1224 questions, 306 in the first phase and 918 in the second. Of the 1224 questions, 745 were answered correctly giving an answer accuracy rate of 61%. Of the correct answers, 388 came from graphical problems (per participant, M=23, SD=8) and 357 came from textual problems (M=21, SD=8). Representation type did not have a significant main effect on the number of correct answers $t(16)$=2.06, p=0.06.

As well as accuracy, we were also interested in the speed at which participants were able to complete the problems. Problems from the first phase were used for this purpose because there was only one question associated with each problem. This allowed us to eliminate any effects of guessing, by examining the time spent on a problem for correct answers only. On average, participants took longer to complete a problem using the textual representation (M=69s, SD=33s) than with the graphical representation (M=48s, SD=20s). This difference was significant[1], $t(15)=2.24$, $p<.05$.

We were unable to meaningfully evaluate the time spent on problems in the second phase, because only accepting problems with three correct answers made for too small a sample. Therefore, the final measure we were interested in was how long it took participants to start answering the first question in a problem. To do this, we measured the time between the presentation of the representation and the first keypress in the answer field. (First click was not used because participants would often click the answer field and then work out the answer.) Including only correct answers, we found that participants took significantly longer to calculate their first answer when given textual representations (M=44s, SD=23s) compared to when they were given graphical representations (M=27s, SD=11s), $t(16)=3.43$, $p<.01$. We do not consider the effect that interruptions had on performance as this is beyond the scope of this article.

The post-experiment questionnaire asked participants to rate their agreement with eight statements about the task. The first three questions in particular were of interest as they asked participants about their subjective sense of the difficulty of the experiment as a whole and the two different kinds of representation. The modal response to the first question "The task I was given was difficult" was 'Agree'. Participants responded 'Neutral' to the second question: "Problems with text representations were harder than problems with graphical representations." There was no single modal response for the third question. "Problems with graphical representations were harder than problems with text representations." The mode was split between 'Disagree' and 'Neutral'.

4 Discussion

The results of the experiment suggest that graphical representations may confer some performance advantages over textual representations when calculating infusion parameters, giving support to previous proposals (e.g. [7]) and confirming that re-representing calculation problems can improve performance (e.g. [10]). Although the effect of representation on answer accuracy was not significant, there was a trend toward superior accuracy in graphical problems. When interpreting this trend it should be noted that participants completed each problem twice – once for each representation – meaning that the difficulty of the problems could not have affected how accurate participants were in their answers.

Stronger evidence for the performance benefits of graphical representations is provided by response time data. These show that participants completed single-answer problems more quickly when presented with graphical representations and answered the first question more quickly in multiple-question problems. Even if participants

1 One participant was excluded from this test because they produced no correct answers to Phase 1 textual problems.

eventually came to the same answer with both representations, they did so more quickly using the graphical representation. These differences in response time support the idea that these representations, while informationally equivalent, were not computationally equivalent; participants' accuracy may have been knowledge-constrained (i.e. by their arithmetic skills), but graphical representations allowed them to operate within these constraints more effectively. In a time-pressured environment, small time savings that have no negative effects on standards of care might prove valuable.

Our results indicate that as well as providing a small increase in accuracy, graphical representations deliver a significant increase in efficiency. Participants' feedback in post-experiment questioning revealed that they felt that graphical representations were somewhat less difficult to use than textual representations; this is despite the fact that they still had to calculate the answer themselves (i.e., they could not just read the answer from the graph). Overall the results of this study demonstrate that graphical representations may confer some advantages to support the calculation of infusion parameters over traditional textual representations of these problems.

In some ways the experiment underutilises the features of graphical representations. For instance, some medications need to be delivered with a loading dose. This requires a high rate for a short period at the beginning of an infusion, followed by a slower infusion for the remainder of the prescription. These kinds of infusion could easily be represented on a graphical representation: the gradient of the line would change to reflect the changing rate of infusion. Conversely, the simple textual representation in this study provides no mechanism for presenting such an infusion in an easily digestable manner. Moreover, graphs were drawn inaccurately so that participants could not compute answers directly from the graph – accurate graphs might have provided more scope for participants to reason about the problems visually. Graphical representations could be embedded into future infusion devices, where they could provide dynamic indication of infusion progress (e.g. [7]), or their benefits could be realised by printing them on medication bags and prescriptions alongside (or instead of) textual representations.

The extent to which graphical representations would be useful on wards hinges on generalisability of the results of this study. We used a lab-based experimental investigation because we wanted to see if representational differences would affect performance in a group of participants who were unfamiliar with either kind of representation; testing the interfaces on nurses will be difficult because they will likely be accustomed to the traditional textual representation. Future work should investigate whether calculation performance improves with graphical representations of infusion by introducing graphical representations to nurse training programmes. Our results, together with previous work which demonstrates that visual representations have advantages over textual representations in pedagogic situations [13], suggest that graphical representations could have positive effects for trainees, where textual representations do not have the familiarity advantage they currently have in hospital wards.

Acknowledgements. Gould, Cox and Brumby are funded by the CHI+MED project, EPSRC grant EP/G059063/1. Thanks to Sarah Wiseman for her comments on this work.

References

1. Bindler, R., Bayne, T.: Medication Calculation Ability of Registered Nurses. Journal of Nursing Scholarship 23, 221–224 (1991)
2. McMullan, M., Jones, R., Lea, S.: Patient safety: numerical skills and drug calculation abilities of nursing students and Registered Nurses. Journal of Advanced Nursing 66, 891–899 (2010)
3. Harvey, S., Murphy, F., Lake, R., Jenkins, L., Cavanna, A., Tait, M.: Diagnosing the problem: Using a tool to identify pre-registration nursing students' mathematical ability. Nurse Education in Practice 10, 119–125 (2010)
4. Schulmeister, L.: Chemotherapy medication errors: descriptions, severity, and contributing factors. Oncol. Nurs. Forum. 26, 1033–1042 (1999)
5. Harne-Britner, S., Kreamer, C.L., Frownfelter, P., Helmuth, A., Lutter, S., Schafer, D.J., Wilson, C.: Improving medication calculation skills of practicing nurses and senior nursing students: a pilot study. J. Nurses Staff Dev. 22, 190–195 (2006)
6. Hutton, M., Coben, D., Hall, C., Rowe, D., Sabin, M., Weeks, K., Woolley, N.: Numeracy for nursing, report of a pilot study to compare outcomes of two practical simulation tools – An online medication dosage assessment and practical assessment in the style of objective structured clinical examination. Nurse Education Today 30, 608–614 (2010)
7. Nemeth, C.P., O'Connor, M., Cook, R.I.: The Infusion Device as a Source of Healthcare Resilience. In: Nemeth, C.P., Hollnagel, E., Dekker, S. (eds.) Preparation and Restoration, p. 215. Ashgate Publishing, Ltd. (2009)
8. Furniss, D., Blandford, A., Mayer, A.: Unremarkable errors: low-level disturbances in infusion pump use. In: Proceedings of the 25th BCS Conference on Human-Computer Interaction, pp. 197–204. BCS, Swinton (2011)
9. Eastwood, K.J., Boyle, M.J., Williams, B., Fairhall, R.: Numeracy skills of nursing students. Nurse Education Today 31, 815–818 (2011)
10. Thimbleby, H., Williams, D.: Using nomograms to reduce harm from clinical calculations. In: IEEE International Conference on Healthcare Informatics (accepted)
11. Larkin, J.H., Simon, H.A.: Why a Diagram is (Sometimes) Worth Ten Thousand Words. Cog. Sci. 11, 65–100 (1987)
12. Zhang, J., Norman, D.A.: Representations in distributed cognitive tasks. Cog. Sci. 18, 87–122 (1994)
13. Ainsworth, S., Th Loizou, A.: The effects of self-explaining when learning with text or diagrams. Cog. Sci. 27, 669–681 (2003)
14. Peebles, D., Cheng, P.C.-H.: Modeling the Effect of Task and Graphical Representation on Response Latency in a Graph Reading Task. J. Hum. Factor Ergon. 45, 28–46 (2003)
15. Shah, P., Freedman, E.G.: Bar and Line Graph Comprehension: An Interaction of Top-Down and Bottom-Up Processes. TopiCS 3, 560–578 (2011)
16. Kirchner, W.K.: Age differences in short-term retention of rapidly changing information. J. Exp. Psych., 1–17 (1958)

Visual Conversational Interfaces to Empower Low-Literacy Users

Sheetal K. Agarwal[1], Jyoti Grover[1], Arun Kumar[1], Monia Puri[1], Meghna Singh[1], and Christian Remy[2]

[1] IBM India Research Lab, ISID Campus, Block C, Vasant Kunj, New Delhi, India-110070
[2] University of Zurich, Binzmühlestrasse 14, 8050 Zurich, Switzerland
{sheetaga,jyogrove,kkarun,monipur,meghna.singh}@in.ibm.com, remy@ifi.uzh.ch

Abstract. Mobile phones have come a long way from being plain voice calling devices to becoming multipurpose handy tools powered by ever increasing new applications available on-the-go. For many, the mobile phone of today has become the essential device one does not leave home without. However, for a large percentage of human population mobile phone apps are not of much use as they are not literate or IT savvy enough to be able to benefit from them. Recent advances in voice-based telecom information systems enable underprivileged and low-literacy users to access and offer online services without requiring expensive devices or specialized technical knowledge. However, voice applications are limited in their capability due to their time consuming nature. In this paper, we demonstrate an interaction modality that combines the power of voice communication with graphical interfaces in smartphones to break the barrier of illiteracy.

Keywords: Diversity, HCI4D, Information Sharing, ICTD, User-Centered Design, Interactive Voice Systems, Smartphones, India.

1 Introduction

Over the last couple of years, smartphones have proliferated at a large scale even in developing countries. Emergence of new local manufacturers has led to easy availability of smartphones in the market costing less than USD 100 [2]. The plethora of mobile apps available today have brought applications in domains ranging from personal management to healthcare, collaboration to education, at users' fingertips. However, even though a lower upfront cost has helped extend the reach of smartphones, their utility still remains limited to voice calls for a large section of their users in developing countries. The primary reason for this is that the running cost of

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 729–736, 2013.

downloading and using several apps is high if an active internet connection is required. Furthermore, another barrier is the low level of literacy or English literacy that prevents many needy users from making use of their powerful handsets effectively.

On the other hand, advances in interaction experience of telephony voice user interfaces have seen good uptake by this underserved population in developing countries [1, 5, 10]. Traditionally, they have not had any access to online information systems due to reasons of affordability, local relevance, and illiteracy.

In this paper, we propose to marry voice user interfaces available on ordinary phones with the power of graphical user interfaces for Visual Conversational interfaces that achieve a two-pronged effect. First one is to enable textually illiterate people to harness the power of mobile apps available on their smartphone devices. Second is to exploit the rich user interface available through mobile apps to make existing applications usable by such users. Examples include even basic device based services such as SMS and address book among others that are currently not usable by low-literate. In this paper, we demonstrate the use of smartphones to enable a voice content sharing telephony application for textually illiterate users without the need for an Internet connection.

2 Voice Interface Intermingled with Visual Interface

To demonstrate the synergy between voice user interfaces and smartphone based graphical user interfaces, we took the scenario of voice content sharing over telephony voice applications. The voice interface part of the application allows callers to identify and generate a link to online voice content of interest. This link is received by the caller as SMS which can then be shared with others. A corresponding mobile application makes use of intuitive icons that let the user call online voice application for accessing voice content. The application also allows navigation of local repository containing links received from others and enables commands for accessing those links without having to open the address book or SMS to punch in the code for those links.

Figure 1 shows screenshots of the application that enables users to share funny messages in a modulated voice, with their friends. On calling the application, the caller is asked to record any content. This content is then voice modulated to resemble a cat voice and is played back to the user. At this point the user can press the *7 key combination to save this recording and generate a link to this recording. If not, the user can simply continue using the application by recording more content. If the *7 key combination is pressed while the modulated content is being played back or a few seconds after it, the application logs the request to save it, generates a link for it and informs the user that he will get the link via an SMS shortly. The received link allows direct access to the stored voice content bypassing any voice menu navigation.

Fig. 1. Screenshots of the Android application: a) home screen, b) list of recordings, c) single recording selected, d) screen during the phone call

To access a link, a feature phone user needs to dial the phone number embedded in the link and enter a numeric code (also embedded in the link) when prompted. When using the smartphone application the process is seamless and it requires only a tap from the user to access the link. On link access, the voice application fetches and plays the corresponding content and continues with the regular voice application interaction (i.e., the caller can record and share more such recordings).

This application enables textually illiterate users to make use of device features such as SMS and address book through the use of an intuitively designed mobile app. Also, it enables mobile apps to take the aid of voice interfaces in local language that these users are comfortable with.

A second example scenario where this concept can be applied is in the context of navigating Interactive Voice Response (IVR) menus of various organizations. Services such as Gethuman[1] and Deepdial[2] provide a mechanism to callers to navigate directly to a particular portion of an IVR's deep navigation menu without having to go through the entire process manually. A mobile application on the user's device could either prefetch or obtain via Short Message Service (SMS) on demand, an outline of the target IVR menu. On connecting a call, it could present a visual interface constructed from that IVR meta data. This would enable the caller to navigate the IVR menu through the visual interface rather than having to speak or punch in a digit everytime. Further, with the concept of voice links, the individual menu options of the IVR could be hyperlinked allowing the user to jump directly to a portion rather than wait for the entire navigation to take place sequentially. Augmenting IVR menus with corresponding visual interfaces was also explored by FonePal system [14] where an Instant Messaging (IM) client was used to present the visual interface.

1 `http://gethuman.com`
2 `http://deepdial.com`

Fig. 2. Researchers demonstrating the visual conversational interface to subjects from target population

3 System Design and Implementation

Even though the processing power and memory available on mobile phones has increased drastically in recent years, yet current speech recognition systems available on mobile devices are not powerful enough to perform sophisticated recognition tasks. Applications such as Siri[3] virtual assistant from Apple and Nina[4] from Nuance make use of cloud based speech recognition software to deliver their service. Since our target users may not have Internet subscription required to utilize such services, our system makes use of ordinary voice channel based interaction with server side voice application deployment and hosting.

To offer these telephony voice services coupled with visual mobile app interface, we make use of applications built over Spoken Web Application Framework [4]. It is a platform that enables easy creation of new voice applications even without the need for any programming skills or even textual literacy. It has been deployed in several applications meant for serving the underprivileged that have language or affordability as a barrier to access or offer information services.

Figure 3 shows a typical deployment infrastructure where a smartphone user without a data connection invokes a mobile application. This mobile application provides its functionality locally and makes a call to a server based voice application deployment. The user could interact with the voice call even as supportive content or icons are displayed on the screen. The server side telephony infrastructure makes use of a Voice Gateway to convert Public Switched Telephone Network (PSTN) calls into voice-over-IP (VoIP). This is then delivered to the application platform which renders Voice XML (VXML) pages to interact with the user.

Nuance Recognizer (or similar system) is used for Speech Recognition through its Automatic Speech Recognition (ASR) function. The database server manages all recordings, content and configuration, while SMS gateway and Email server provide a

---

[3] `http://www.apple.com/ios/siri/`

[4] `http://www.nuance.com/meet-nina/index.htm`

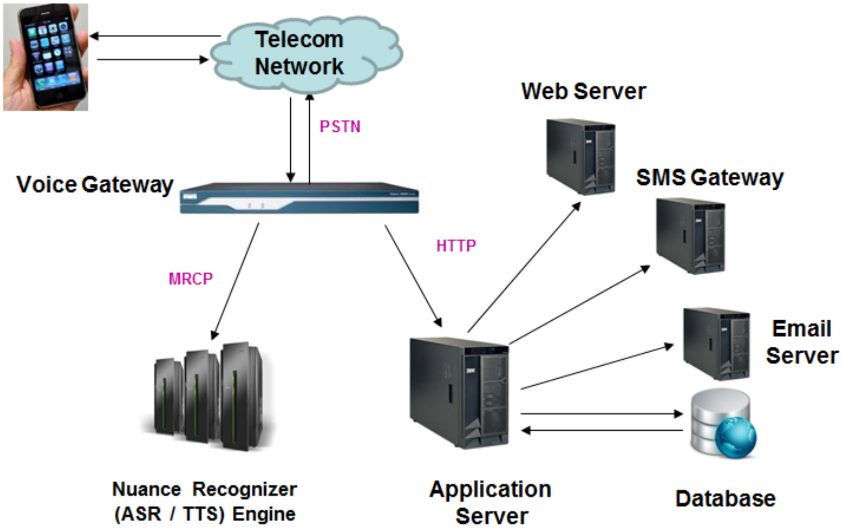


Fig. 3. Infrastructure setup for providing conversational interface to mobile applications

mechanism to send and/or receives SMS and email respectively. The mobile application can interact with server on any of these channels – voice, SMS, email. A web server is shown since in some scenarios, the stakeholders may like to see the reports and other function through a web based interface.

Since the voice interface is available over ordinary telephone call, it does not requiring any sophisticated device capabilities at the client's end.

4 Related Work

The concept of combining a conversational interface with a visual one has been applied in the context of visualization tools [11]. To create effective visualizations the Articulate system provides a semi-automated visual analytic model coupled with a conversational user interface. Using natural language processing and some heuristics it tries to create a suitable visualization as desired by the user.

Employing conversational interfaces for software agents was employed as early as in the 90s [12] and some of those concepts are visible in intelligent automated assistant systems of today such as CoCo system [6] that uses conversational interface to automate tasks on the web as well as Siri and Nina. They are also being applied to other tasks such as information retrieval [5].

Interplay of a conversational interface with animated personas has been studied by Oviatt et al. [7]. In their study with children, they learnt that children's speech converged with the text-to-speech (TTS) heard from the animated persona. The participating children adapted several acoustic-prosodic features of their speech based upon what they heard from animated personas. Further, children readapted when exposed to new personas.

However, the focus for such conversational systems primarily has been to enable multiple channels of communication with a software system or to study the influence of multiple channels of communications. Recent popular tools such as Siri or Nina, focus primarily on the conversational aspect of speech based interface and do not attempt to leverage the conversation interface with a corresponding visual interface.

This paper, on the other hand, proposed an interaction modality in which applications leverage conversational interfaces side by side and intermixed with visual interfaces to help several users overcome their accessibility problem. These could be textually-illiterate people or older adults unable to read or type properly especially on mobile phones. An entertainment application with such a visual conversation interface has been presented in [9].

5 Discussion

Due to the increased proliferation of smartphones, it is only a logical step to leverage the power of visual interfaces and provide illiterates with additional means of communication. Friscira et al. [3] used a smartphone application to augment SMS messages with icons that let illiterate users make use of basic SMS messaging functionality. This marks just the beginning of a range of new possibilities that open up for researchers and practitioners. As we found out in a study focusing on the aspect of sharing information [9], smartphone applications can contribute to the understanding of interactive voice application services and facilitate the interaction with such voice applications.

The example content sharing implementation we presented in this paper could be extended to allow for sharing to multiple contacts simultaneously by simply clicking on pictures of contacts stored in the address book. This would not only simplify sharing, it would also save time and therefore money for the caller of such a service, and make the process less error-prone by removing the need to enter all recipients' numbers via phone keyboard or voice input.

An area of growing interest in research in developing countries is that of job opportunities for the underprivileged [8, 13]. Browsing such jobs by voice navigation only is cumbersome and time-consuming, eventually reducing effectiveness and success of such applications. Not only does this make such services more useful, the visualization might also contribute to the understanding of the hierarchical structure of voice systems.

6 Conclusion

We presented a new approach of creating visual conversational interfaces that utilize the voice interface of telephony voice applications along with rich graphical interfaces of smartphones to help low literate users overcome their handicap. While providing this, the application does not assume the presence of a data connection on the phone. In the presence of a data connection, these interfaces can be made a lot richer and capable than what is possible with just voice connectivity and local processing power. This mode of interaction (with or without data connection) is also applicable to other users with accessibility challenges such as older adults. As mobile phone platforms become more powerful to host voice applications locally or Internet usage charges become affordable for this population, voice applications could be much more seamlessly integrated with mobile applications similar to emerging voice-on-cloud based mobile assistant applications of today.

References

1. Agarwal, S., Kumar, A., Nanavati, A.A., Rajput, N.: Content creation and dissemination by-and-for users in rural areas. In: Proceedings of International Conference on Information and Communication Technologies and Development (ICTD), Doha, Qatar (2009)
2. Datta, A.: "New low-cost smartphone from Micromax". The Hindu Business Line, `http://www.thehindubusinessline.com/industry-and-economy/info-tech/new-lowcost-smartphone-from-micromax/article3357066.ece` (last accessed March 2013)
3. Friscira, E., Knoche, H., Huang, J.: Getting in touch with text: designing a mobile phone application for illiterate users to harness SMS. In: Proceedings of ACM Second Annual Symposium on Computing for Development (DEV), Atlanta, Georgia (2012)
4. Kumar, A., Agarwal, S., Manwani, P.: The spoken web application framework: user generated content and service creation through low-end mobiles. In: Proceedings of the International Cross Disciplinary Conference on Web Accessibility (W4A), Raleigh, USA (2010)
5. Lau, T., Cerruti, J., Dixon, M., Nichols, J.: Towards conversational interfaces to web applications. In: Proceedings of Fifth Workshop on Human-Computer Interaction and Information Retrieval (HCIR), Mountain View, CA (October 2011)
6. Lau, T., Cerruti, J., Manzato, G., Bengualid, M., Bigham, J.P., Nichols, J.: A Conversational Interface to Web Automation. In: Proceedings of 23rd ACM Symposium on User Interface Software and Technology (UIST), NY, USA (2010)
7. Oviatt, S., Darves, C., Coulston, R.: Toward adaptive conversational interfaces: Modeling speech convergence with animated personas. ACM Transaction on Computer-Human Interaction 11(3), 300–328 (2004)
8. Raza, A.A., Haw, F., Tariq, Z., Pervaiz, M., Razaq, S., Saif, U., Rosenfeld, R.: Spread and sustainability: the geography and economics of speech-based services. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Paris, France (2013)
9. Remy, C., Agarwal, S., Kumar, A., Srivastava, S.: Supporting Voice Content Sharing among Underprivileged People in Urban India. To appear in Proceedings of 14th IFIP TC13 Conference on Human-Computer Interaction. Cape Town, South Africa (September 2013)
10. Sambasivan, N., Weber, J.S., Cutrell, E.: Designing a Phone Broadcasting System for Urban Sex Workers in India. In: Proceedings of ACM CHI Conference on Human Factors in Computing Systems (CHI), Vancouver, British Columbia (2011)
11. Sun, Y., Leigh, J., Johnson, A.E., Chau, D.: Articulate: a Conversational Interface for Visual Analytics. In: Proceedings of the IEEE Symposium on Visual Analytics Science and Technology, Atlantic City, New Jersey (2009)

12. Trower, T.: Creating Conversational Interfaces for Interactive Software Agents. Tutorial Proceedings of ACM CHI Conference on Human Factors in Computing Systems (CHI) (1997)
13. White, J., Duggirala, M., Srivastava, S., Kummamuru, K.: Designing a Voice-based Employment Exchange for Rural India. In: Proceedings of International Conference on Information and Communication Technologies and Development (ICTD), Atlanta, Georgia (2012)
14. Yin, M., Zhai, S.: The benefits of augmenting telephone voice menu navigation with visual browsing and search. In: Proceedings of ACM Conference on Human Factors in Computing Systems (CHI) (2006)

Visual Indication while Sharing Items from a Private 3D Portal Room UI to Public Virtual Environments

Minna Pakanen[1], Leena Arhippainen[1],
Jukka H. Vatjus-Anttila[1], and Olli-Pekka Pakanen[2]

[1] Intel and Nokia Joint Innovation Center, Center for Internet Excellence,
P.O. Box 1001, FI-90014 University of Oulu, Finland
[2] University of Oulu, Department of Information Processing Science,
P.O. Box 3000, FI-90014 University of Oulu, Finland
{minna.pakanen,leena.arhippainen,jukka.vatjus-anttila}@cie.fi,
opakanen@paju.oulu.fi

Abstract. In this paper, we describe the user experience evaluation results of a 3D Portal Room UI for sharing 3D objects from private space to remote public virtual environments. The user evaluation with 30 participants was conducted with a functional prototype and additional high quality images that were printed on paper sheets. The evaluation indicated that participants liked this way of sharing objects and found it also useful. However, it also raised some privacy concerns, especially if the target virtual environment was perceived as public. Evaluation elicited that the visual indication while sharing objects is important; therefore, designers of 3D virtual environments should prefer a distinguishable glow around the shared object and portal.

Keywords: Portals, 3D user interface, virtual environment, user interaction, visual indication, user experience.

1 Introduction

In collaborative three dimensional (3D) virtual environments (VEs), such as Second Life [9] and World of Warcraft [2], users can see objects and other people's avatars in a 3D space. To share items with other users or players, current solutions must rely on two dimensional (2D) pop-up repository windows, which reduce the view of the VE and weaken the 3D experience. There is a need for a private 3D VE which is located in parallel with public VEs and the possibility to move objects between them. The need is noted in collaborative VE context, where a user wants to have private information in collaborative VE, but at the same time, he/she needs to be able to hide it form other users [3]. There is also a need in education context, where a teacher could prepare classes and share material from his/her private 3D space to the pupils in public virtual classrooms, and at the same time, be able to monitor what pupils are doing in collaborative VEs [1].

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 737–744, 2013.

In this paper, we present a 3D Portal Room UI with an ability to have four client server connections in parallel, which are presented to the users as four parallel views to the VEs. The focus of this paper is on visual indication when moving 3D objects from a private 3D Portal Room UI to public VEs. We report the user evaluation of the prototype and additional visual indication image examples with 30 participants. This paper contributes to the 3D VE research and the designer community by presenting user experiences (UXs) [4] with the prototype and participants preferences on visual indication while objects are shared through portals.

2 Related Research

Portals have been studied in prior research from the navigation point of view [7]. Kotziampasis et al. [7] found that portals help users to navigate in the virtual worlds, because of the visual presentation of the target space. Portals have also been used as object transfer devices, but only in two dimensional environments, where documents and 2D shapes can be exchanged through portals on a surface display [15].

The limitation of prior research is the lack of user interaction research. Shneiderman [14] argued ten years ago about making 3D UIs to facilitate user tasks with an enhanced 3D design rather than just mimicking reality in all possible ways. To design visually enhanced 3D UIs and games, we agree with El-Nasr and Yan that visual indication is an important factor, especially for novice users [5].

Visual attention process has been studied extensively within cognitive science and psychology, but not from a 3D game perspective [5]. Visual indication in 3D scenes has been studied from the anchor highlighting point of view, where the colour coding on 3D objects was preferred in the usability test and user comments, but the most visually appealing choice was the colour edges [11]. Also, it is noted in 3D games research that the location of the visual indication is as important as the visual indication itself [5].

Butz et al. [3] have studied indication of shared elements in collaborative 3D VEs from privacy perspectives. They introduced two methods for privacy management: a vampire mirror, which shows only shared items, and a privacy lamp, in which a beam of light depicts the private items of other objects in collaborative 3D VE in augmented reality situation [3]. Also, it is suggested in prior research that the colour of the shared element should be changed or other visual feedback be given when something is moved from a personal 3D GUI to the collaborative 3D VE [10].

3 The Evaluation Prototype

We used the realXtend Tundra 2 [12] virtual world viewer as a setup platform for portals. It is based on a client-server type architecture. We used the following hardware setup in the user evaluation: HP Elitebook 2760p laptop with Ubuntu Linux 12.04 LTS 32bit with WLAN 802.11g public Internet connection. In addition, we had an external 24” Dell P2411Hb 1920x1080 monitor, mouse and keyboard.

The 3D Portal Room UI acts as a private 'log-in screen' to the other VEs. Portals provide an ability to have four client-server connections in parallel, which are presented to the users as four parallel views to the VEs [6]. We decided to use a room metaphor, because it creates a feeling of privacy [10]. As we wanted to increase the feeling of private space, there are no windows in the room and the wall texture was chosen to look sturdy (Fig.1). In the 3D Portal Room UI, there are the user's personal items such as a microphone, a photo album, a notebook, a trashcan, a game and MS PowerPoint, Word, PDF files. The portals (views) to the public VEs are covered with round doors, when there is no open connection to them. Connection is created by tapping the door, when a door mesh is replaced by a real-time view to the remote VE. We chose four remote public VEs (Music Club, Office, Outdoor Music Club and City) for the prototype. The chosen VEs had different visual styles from each other, because we wanted to study how participants perceive sharing items to them. (Fig. 1.)

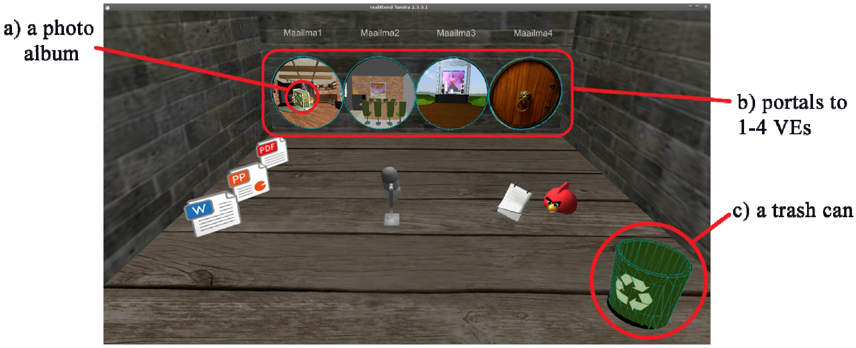


Fig. 1. In the 3D Portal Room UI, the visual indication is showed on the selected object (a) and on possible targets (b, c) when a user is sharing a photo album to the Music Club VE

We implemented to the 3D Portal Room UI a functionality to move private 3D objects to remote VEs and destroy objects from the private UI by moving them to a 3D trashcan. The visual indication for the selected object and targets was implemented with turquoise lines and green overlay colour on the object's mesh (Fig. 1).

The idea was to implement the prototype on a touch screen device, but due to limitations of the Linux Ubuntu operating system, it was not a suitable solution. Therefore, we decided to evaluate the prototype with mouse input to get user feedback on the portal idea of moving items through portals and the visual indication while doing it. The main idea with implemented interactions was that a user selects an object by pressing the left mouse button, and then he/she moves the cursor on the VE portal, when an object 'jumps' on the portal. The user then presses the left mouse button and a copy of the object drops on the ground into the VE and the original object returns to its place in the 3D Portal Room UI. We also implemented realistic physics for the shared object when it drops into the VE. Thus, we did not define the position where the shared object should land in the VE; therefore, the objects landed with a random place and orientation.

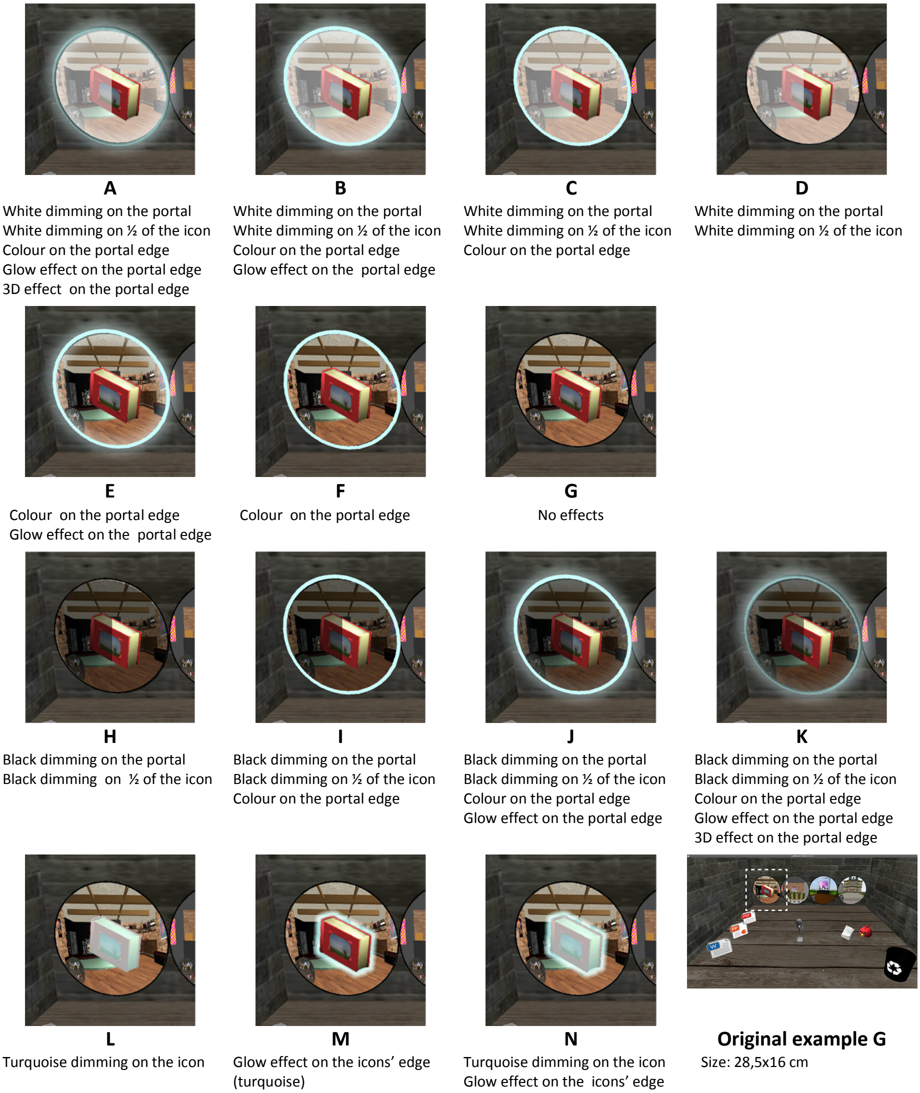


Fig. 2. Enlargement of the visual indication effects used in the example images (A-N). Below an example of the original image with the enlargement area indicated with dashed white line

4 Example Images for Evaluation

To study what kind of visual indication users would prefer for indicating moving objects from their private 3D UI area to public VE, we prepared fourteen different kinds of indications either on the object or on the target portal (Fig. 2). We used as an inspiration object highlights that we had noticed in games, such as Serious Sam 3 BFE [13]. We wanted to extend the visual indications used in prior research [11] and investigate suggested solutions [10]; therefore, we made several versions with different mixtures of colour overlay, colour edges and glow. We decided to use a uniform colour theme in all options, to make them easily comparable with each other. Indications were made in the Adobe Photoshop by adding different kinds of layer styles for the screenshot image of the 3D Portal Room UI. We decided to present visual indication examples as images printed on 28,5x16cm sized paper, because then participants can compare examples easier by taking them into their hands. It is also a good way for collecting user experiences in an early development phase and it is also more cost-effective and faster than implementing all possible solutions to the prototype [10].

5 User Evaluation

We recruited 30 participants for the user evaluation. The participants' age varied from 20 to 52 years with a mean of 32. 66% were male. 60% of the participants used 3D VEs on a computer, but none on a tablet or a mobile phone.

A user evaluation had two parts. In the first part, participants familiarized themselves with the prototype by doing simple tasks with it. In the second part, subjects ranked three best options from given examples (A-N) to represent visual indication while an object is moved from the private 3D UI to a public VE. Subjects were interviewed and their subjective experiences were gathered by observing, writing down their comments and video recording while completing the following tasks:

1. Open connections to the VEs (click doors) & explain if the VE is public or private.
2. Share objects from the private 3D Portal Room UI to public virtual environments:
   (a) Share Power Point slides to the Office scene (drag PP icon to VE 2)
   (b) Give feedback to the City council (City scene) (drag W icon to VE 4)
   (c) Delete an old PDF file (drag PDF icon to the trashcan in 3D Portal Room UI)
   (d) Share a photo album to the Music Club (drag photo album 3D icon to VE 1).
   Comment on the following question: Do you find sharing (1) easy, (2) useful, (3) secure, and (4) what kind of indication there should be presented while sharing.
3. Select from given alternatives (A-N) the best, 2nd best and 3rd best option for indicating an object moving through the portal to a public VE. The images were placed on the table in random order for each subject. Participants were asked to mark their choice with little Post-it notes and comment each selection.

6 Findings

For all of the participants, it was easy to understand the logic that by clicking the door a connection opens to the VE. When sharing the objects, 70% of the users interacted directly as intended and 27% did it correctly in the second try. 3% did it right in the third try. The main problem with the interaction during sharing was that participants did not press the left mouse button on the portal, to drop a copy of the object to the VE. As one user commented: *"Why I have to press* (left button) *-it is weird... I would just drag it there"*. Also the object transition implemented in the prototype caused confusion for a few participants when they were sharing objects. As they moved the cursor slowly, it seemed that the selected object did not follow the cursor at all. As a consequence, users thought that they have not selected the object at all. As one user stated: *"*(object) *should follow* (the cursor) *smoothly and not just jump on the portal"*.

6.1 Security Concerns while Sharing Objects

In general, participants liked the possibility to share items to the remote VEs and they also found it useful, especially with the case of sharing a PowerPoint file to the Office. However, sharing also raised privacy and security concerns, especially with VEs that participants thought to be open for all people. All of the participants thought that outdoor VEs (Outdoor Music Club and 3D City) were public. Also the indoor Music Club was perceived to be public by 97% of the participants. The Office was perceived public only by 57% of the users. Participants explained that they perceived it either private or semi-private because the environment was quite small and the amount of the chairs in the scene created a feeling that it is meant for a limited amount of attendees. Subjects explained also that in real-life, meeting rooms are meant for private gatherings.

6.2 User Experiences with the Visual Indication in the Prototype

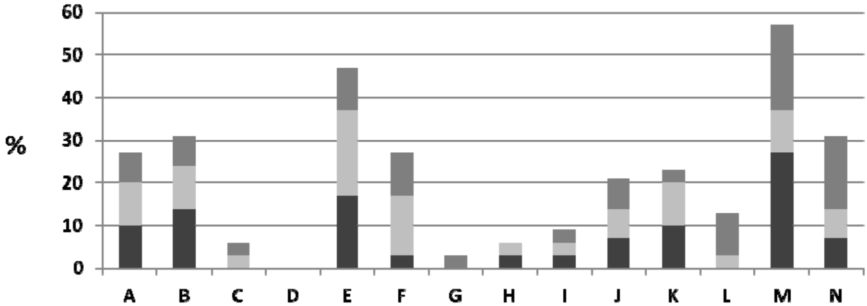
Participants liked the realistic physics when a shared object was dropped to the virtual environment. It was found as important but also amusing. They commented it as: *"Wow"*, *"Funny" and "visually attractive"*. One participant commented: *"Dropping is a good effect indicating that* (an object) *goes to the virtual environment"*. Participants did not like the fact that the objects did not land in a rational location in the scene and they wished to be able to organize and move the objects in the VE after sharing them.

The implemented indication while sharing was not perceived to be visible enough. Participants commented that they wished to have clearer visual feedback on the edge of the shared item or/and on the portal, such as *"light"*, *"flash of light"* or *"colour change"*. Also audio feedback was recommended. What is more, a few wished for a check-up window with text: *"do you want to share this item to the ..."* to pop-up.

6.3 User Experiences with the Visual Indication Example Images

The distribution of participants' preferred visual indications is presented in Figure 3. To weight the primary choices, we used Borda count [8]. We gave three points for participants' 1st choice, two points for their 2nd choice, one point to the 3rd choice and zero points to all other options.

A non-parametric Kruskall-Wallis test indicated significant differences between the data points for the fourteen different alternatives ($\chi 2(13) = 27{,}530$, $p < 0{,}02$). In the preference ranking task, participants preferred option M (turquoise glow effect around the edge of the photo album) for indicating objects transferring from private 3D UI to the public VE. It received the highest amount in the Borda count (36) and 27% of the participants rated it as the best choice. Participants explained that the



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
The Best	10%	14%	0%	0%	17%	3%	0%	3%	3%	7%	10%	0%	27%	7%
2nd Best	10%	10%	3%	0%	20%	14%	0%	3%	3%	7%	10%	3%	10%	7%
3rd Best	7%	7%	3%	0%	10%	10%	3%	0%	3%	7%	3%	10%	20%	17%

Fig. 3. Preference ranking results for the visual indication

indication of the object is the most important thing to be represented. They also commented that when the visual indication is on the object it looks like something is happening; the object is moving to the other VE. The glow effect was also perceived *as "visually pleasant"*, *"easily noticeable"* and *"it really makes the object to pop-up"*.

The second highest was option E (turquoise colour + glow effect around the edge of the portal). It received 30 points and 17% of the first-place votes. Participants explained that the most important thing is now highlighted (e.g. the target). As one participant commented: *"I have already chosen the object, so the glow indicates where to drop it"*. It was also suggested that both the object and the portal could glow.

The third highest was option B (white dimming on the portal view and on half of the photo album, turquoise colour + glow effect around the edge of the portal) with 20 points and 14% of the first-place votes. It was preferred because it gave an impression that the object is already half way through to the other VE, therefore, it looked intuitive. Also the indication on the portal was large enough to be noticed, as one participant commented: "(the indication should be) *as visible as possible to preventing unintentional sharing to the other virtual environment"*. Despite different user views, a pairwise comparison between the three most popular alternatives revealed no significant statistical difference between them.

7 Conclusion

In this paper, we present a private 3D Portal Room UI for sharing 3D objects to public VEs. We present user evaluation with 30 participants and results on both; interaction with a functional prototype and a preference ranking of high quality example images. The participants found the sharing of objects from a private 3D UI space to public VEs as easy and useful, especially when sharing a PowerPoint file to the Office scene. The visual indication should be clear when an object is moved out of the private UI area. The evaluation indicated that designers of 3D VEs should prefer a distinguishable glow effect around the shared object and the target portal.

Whereas we believe that our results provide useful information for researchers and designers of the 3D VEs, we also recognize the limitations of the study. As we decided to use printed paper sheets as examples in the preference ranking task, we missed the interaction with the examples. However, the method provided us a lot of UX based information in the early development phase and was cost-effective and time-saving. With the feedback, we now know how to proceed with the design and implementation of portals. In the future, we are going to study how the visual indication works with the touch interaction. It would also be interesting to study how shared objects should be indicated to the users of public VEs.

References

1. Arhippainen, L., Pakanen, M., Hickey, S., Mattila, P.: User Experiences of 3D Virtual Learning Environment. In: 15th International Academic MindTrek Conference, pp. 222–227. ACM Press, New York (2011)
2. Blizzard Entertainment Inc. World of Warcraft, `http://us.battle.net/wow/en/`
3. Butz, A., Beshers, C., Feiner, S.: Of Vampire Mirrors and Privacy Lamps: Privacy Management in Multi-User Augmented Environments. In: 11th Annual Symposium on User Interface Software and Technology, pp. 171–172. ACM Press, New York (1998)
4. ISO DIS 9241-210:2010. Ergonomics of human system interaction - Part 210: Human-centred design for interactive systems. International Standardization Organization (ISO). Switzerland
5. El-Nasr, M.S., Yan, S.: Visual Attention in 3D Video Games. In: Advances in Computer Entertainment Technology, Article No. 22. ACM Press, New York (2006)
6. Hickey, S., Arhippainen, L., Vatjus-Anttila, J.H., Pakanen, M.: User Experience Study of Concurrent Virtual Environments with 2D Tab and 3D Portal UIs. In: 19th International Conference on Engineering, Technology and Innovation, ICE, June 24-26 (2013)
7. Kotziampasis, I., Sidwell, N., Chalmers, A.: Seamlessly Integrated Distributed Shared Virtual Environments. In: 20th Spring Conference on Computer Graphics, pp. 138–147. ACM Press, New York (2004)
8. The mathematics of voting: Democratic symmetry. The Economist, p. 83 (March 2000)
9. Linden Lab. Second Life, `http://secondlife.com/`
10. Pakanen, M., Arhippainen, L., Hickey, S.: Studying Four 3D GUI Metaphors in Virtual Environment in Tablet Context. In: 6th International Conference on Advances in Computer-Human Interactions, pp. 41–46. ThinkMind Press (2013)
11. Pichler, M.: Interactive Browsing of 3D Scenes in Hypermedia: The Hyper-G 3D Viewer, `http://ftp.iicm.tugraz.at/pub/theses/mpichler.pdf`
12. realXtend, `http://realxtend.org/`
13. Serious Sam 3 BFE, `http://store.steampowered.com/app/41070/`
14. Shneiderman, B.: Why Not Make Interfaces Better than 3D Reality? IEEE Computer Graphics and Applications 23(6), 12–15 (2003)
15. Voelker, S., Weiss, M., Wacharamanotham, C., Borchers, J.: Dynamic Portals: a Lightweight Metaphor for Fast Object Transfer on Interactive Surfaces. In: ACM International Conference on Interactive Tabletops and Surfaces, pp. 158–161. ACM Press, New York (2011)

Web Design for Science Museum towards Engaging User Experience

Mohd Syaheezam Asyraq Yamin and Eswati Azni Jaafar

MIMOS Berhad, Technology Park Malaysia, Bukit Jalil,
Kuala Lumpur, Malaysia

Abstract. Nowadays, exhibition experience is no longer defined by physical visits. Increasing a person's knowledge about a subject tends to increase their interest in it, thus improve attendance and support for the exhibition. This research analyses and conduct comparison studies regarding website elements featured in 30 science museums around the world. Consequently, this produces an overview of one effective model for Website design; a user-centered process that includes techniques for need assessment, methodology, goal/task analysis, user interface design, and finally pre/post prototyping.

Keywords: User Experience, Usability, Web Design, Science Museum.

1 Introduction

As the Internet continues to grow as main sources of information, the design of effective Websites becomes increasingly important. In Malaysia, the level of awareness on issues such as user interfaces and effectiveness regarding website has increased. Some of them even regularly improvise and update their website. However the results are still far from satisfying. Based on the report by "Malaysia Government Portals and Websites Assessment (MGPWA) 2011", only 285 out of the 1,155 websites manage to meet 5-stars rating set by Multimedia Development Corporation (MDEC), a body that oversees Malaysia's Information and Communication Technology initiative. In order to get a 5-star rating, several criteria were put together in MPWGA mostly related to best practice and global standards. This problem also significantly affected other institutions like the Museum and Science Centre. Website is critical as it provides the first impression of an institution. In normal circumstances, Website with great visual design and aesthetic will give more credibility. This report examines the effects of good user interface design for Science Centre Website and whether the implication of user experience design will contribute to number of actual physical visits to the Science Centre.

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 745–754, 2013.

2 Background

Nielsen (2000) stressed that there are two basic approaches to web design, which are: the ‘artistic ideal’ that shows the designer’s self-expression and the ‘engineering ideal’ that delivers solutions to users. Based on that, 3 main areas are considered:

2.1 User Experience Issues

Badly designed websites frustrate users and cause them to leave, as they cannot find the required information. Most of them have bad experience when navigating through the website. Reasons cited for the users’ negative experience include outdated information, difficult to use and mostly, not finding what they wanted (Nua, 2002). To encourage users to return, website should be designed to facilitate users in finding what they need and increase satisfaction while accomplishing their tasks (N. Abdelmessih, 2001). The website should be able to provide an engaging experience to visitors.

2.2 Design Principle

In order to produce a remarkable design, web designer used to manipulate all elements of design and graphic art. Some of these factors range from elements of space, use and size of images, use of animation, colours, audio and many more. A study by Shenkman and Jonsson (2000) and Tractinsky et al. (2000) suggests that visual appearance is important in users’ preference for website. In similar vein (Tractinsky et al. (2000)) stated that ‘what is beautiful is usable’. The design of visual presentation of web pages is based on an imaginative stimulus that, in the intentions of the designer, should represent the best way to implement this strategy. (Visciola, 2000). In order to develop a good visual presentation, designers should be able adapt the design principles to the required design.

2.3 Usability

According to ISO 9241-11 (1992) usability is defined as the ‘extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use’. Most studies agreed that the measurement of effectiveness of the site normally reflected by the satisfaction level of the users expectations, not the researchers and designers. The quality of user–site interaction is not totally depending on the users’ opinion about a website, but also the perception of the identity of the organization providing it, and as well as the evaluation of its efficiency. In this context, each page element has its influence where strategies for access, navigation and orientation are very important aspects. (Mario, 2004)

3 Method

The authors aimed to investigate two issues. Firstly, to find out the main Design and Multimedia elements that could provide good user experience for a science museum website. Secondly, to explore processes that can provide good website experience. In conducting this research, there are several methodologies employed. The methodological approach in this research is a combination between qualitative and quantitative research as depicted in Table 1 and Table 2 below:

Table 1. The qualitative method used to clarify the issues

Purpose	Method	Justification
1.To understand the objectives and goals of PSN	*Unstructured Interview*	To gain comprehensive Information about targeted type of users. A set of questions were emailed to PSN, followed by interview session with their personnel.
2.To explore the experience of site-visitors.	*Observation*	Visual observation of site visitors. Visitors are asked to browse the PSN website and perform a few simple tasks.

Table 2. The quantitative method used to clarify the issues

Purpose	Method	Justification
1. To generalise groups of people.	*Demographics Research*	To identify the potential users of the website.
2. To get more info on the website users (primary data)	*Questionnaires*	To get the right direction towards the development.
3. To compare pattern matching trends, gaps and features	*Visual Analysis*	30 science museums around the world has been analysed based on Design Principles, Usability and Aesthetics factors

4 Results

In any web development process, the most important step is to understand the users. This is the highest priority during any web design project. Findings is clearly defined below:

4.1 Demographic Findings

Demographic survey as depicted in Figure 1 showed the number of visitors that visit Pusat Sains Negara (PSN) or National Science Centre in 2011. The survey reveal that most of the visitors come from adults and children category that later can be simplify into 'family category'. Based on this fact, both PSN and researchers agreed to focus on family with children as the main target audience.

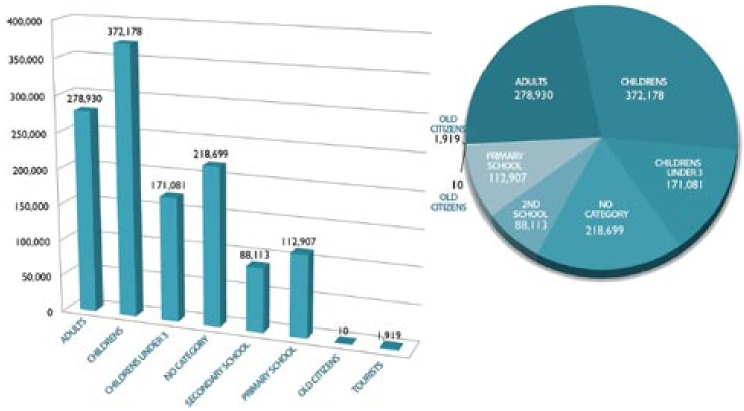


Fig. 1. Demographics analysis for PSN visitors

4.2 Online Surveys

According to the surveys, 50% of the participants state, "planning a visit to Pusat Sains Negara" as their main purpose of visiting the PSN website. This is important finding, since their decision to proceed with the visits might depend on the website credibility itself. The participants were then asked to evaluate the current PSN website based on 4 main categories – Design, Features, Ease of Use and Errors Prevention. As expected, the total score of the website is 2.5 out of 5.The survey clearly indicate that the PSN website need improvements, in terms of design, layout and also presentation.

4.3 Visual Analysis

For Visual Analysis, the first method use is Screen Real Analysis, which is based on 'Homepage Usability: 50 Websites Deconstructed' (Neilsen And Tahir, 2002). As a

Table 3. List of Science Centers

No	Science Museum	*URL*
1	Citi des Sciences et de l'Industrie	*http://www.cite-sciences.fr/fr/cite-des-sciences/*
2	Science Museum, London	*http://www.sciencemuseum.org.uk/*
3	Shanghai Science and Technology Museum	*http://www.sstm.org.cn*
4	National Science and Technology Museum Taiwan	*http://www.nstm.gov.tw/english/*
5	Museum of Science and Industry, Chicago	*http://www.msichicago.org*
6	Pacific Science Center, Seattle	*http://www.pacificsciencecenter.org*
7	Museum of Science, Boston	*http://www.mos.org*
8	Science City, Kolkata	*http://www.sciencecitykolkata.org.in*
9	Ontario Science Center	*http://www.ontariosciencecentre.ca/*
10	Deutsches Museum, Munich	*http://www.deutsches-museum.de/*
11	California Science Center Los Angeles, Los Angeles	*http://www.californiasciencecenter.org/*
12	Scientific Center Kuwait	*http://www.tsck.org.kw/*
13	Orlando Science Centre	*http://www.osc.org/*
14	Maryland Science Center	*http://www.mdsci.org/*
15	Science Center	*http://www.sciencecenter.org/*
16	Petrosains	*http://www.petrosains.com.my*
17	Exploratorium	*http://www.exploratorium.edu/*
18	Scitech	*http://www.scitech.org.au/*
19	Chicago Children's Museum	*http://www.chicagochildrensmuseum.org*
20	Children's Discovery Museum	*http://www.childrensdiscoverymuseum.net/*
21	Children's Museum	*http://www.cmhouston.org*
22	Children's Museum	*http://lsc.org/*
23	Liberty Science Center	*http://www.pleasetouchmuseum.org/*
24	Please Touch Museum	*http://www.childrensmuseum.org/*
25	Children's Museum Indianapolis	*http://www.nhm.ac.uk/*
26	Natural History Museum	*http://www.cdm.org/*
27	Children's Discovery Museum	*http://www2.fi.edu/*
28	The Franklin Institute	*http://www.calacademy.org/*
29	California Academy Of Science	*http://www.puppet.org/*
30	Center for Puppetry Art	*http://www.brooklynkids.org/*

benchmark and comparative studies, 30 science museums around the world has been selected as depicted in Table 3. This museum were selected based on recommendation from Museum Planning website (http://museumplanner.org/worlds-top-10-science-centers/) and from PSN benchmark.

One of the most important aspect to compare is pattern matching – trends, gaps and features. All 'estate' in the websites were scanned properly and then being colour-coded one at a time. Results as depicted in figure 2 below:

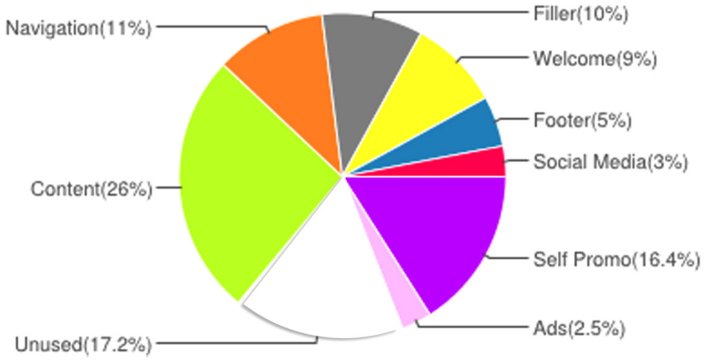


Fig. 2. Summary of screen real estate for 30 museum websites

Visual Analysis on the 30 science centers websites shows that most of the websites have a contrast point (visual differentiation between two or more elements) and also good balance (overall distribution of the visual weight), but most of them suffer badly when it comes to the flow principle which is (the path the users' eyes take through the design). Most of the design failed to combine the design elements in proper manner.

5 Design and Development

Before starting any design process, it is important to establish a clear direction on the new website. Based on the discussion with PSN and the websites comparison depicted in Table 1, two main attributes are derived based on these keywords:

Table 4. Keywords based on feedback from PSN

ATTRIBUTE	KEYWORDS
Corporate	*Accurate, government agency, formality, MPWGA guidelines, Services*
Playful	*Imagination, exploration, curiosity, enjoyment, fun, wow factor, futuristic*

Based on the two attributes above, a design matrix has been composed that is illustrated in Figure 3 below:

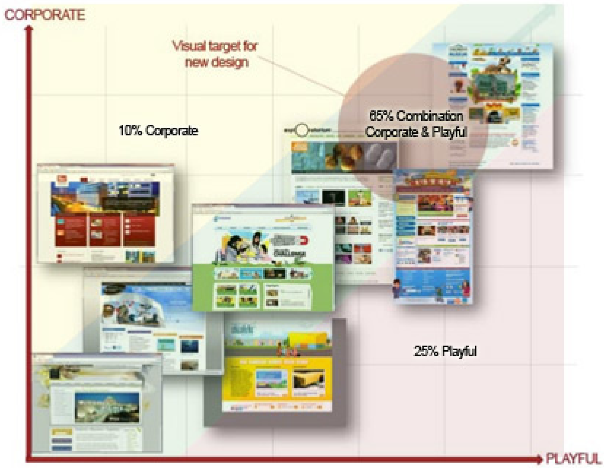


Fig. 3. Design Matrix for the new website

For the development of a new website, *scrolling innovation* has been implemented. It enables user to move some of the graphic elements in different way, or change its size, colour or content whenever user scroll down the website. Another effect implemented in order to enhance user experience is *parallax effect*. It is a technique that featured layered images that moves around the website in different perspectives, which resulting in a nice and interesting 3D illusion. This effect can be achieved using jQuery scripting, CSS3 and HTML5.

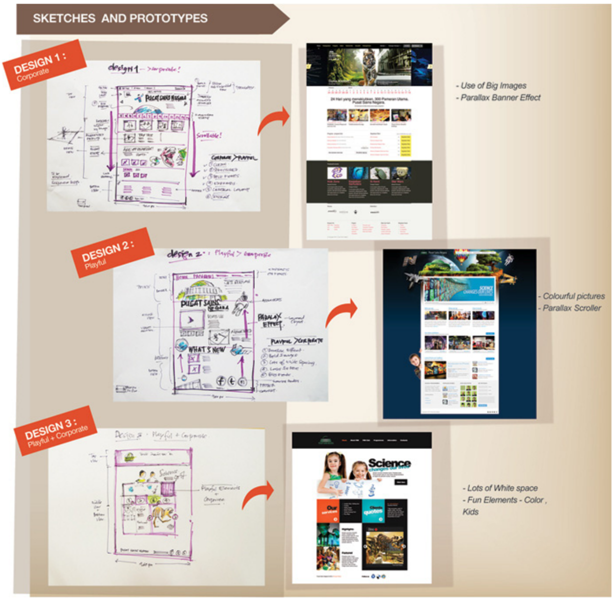


Fig. 4. Sketches and prototypes

Fig. 5. Post Test at Usability Lab MIMOS

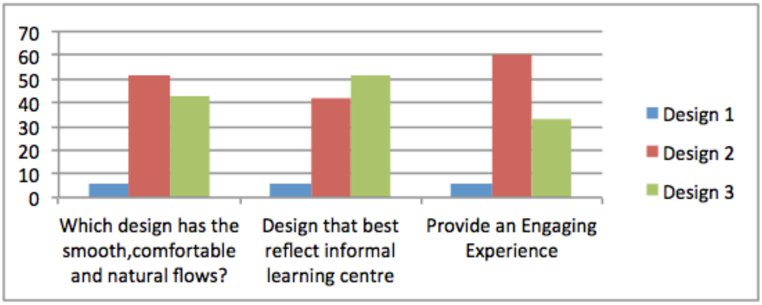


Fig. 6. Analysis for user acceptance of prototypes

Before starting with the actual design development, it is important to start with simple sketches. Researchers has decided to come out with 3 forms of prototypes (Design1, Design 2 and Design 3) as illustrated in figure 4.

A pre and post-test is then conducted before proceeding with the final prototype. Pre-test analysis implies that 'Design 2' is preferred as the most smooth and comfortable design. 62% of the participants also agreed with the use of 'scrolling innovation' while 85% believed that the use of parallax effect has improved their experience as stated in Figure 6.

6 Result

Research findings conclude that it is best to define the elements based on Design Principles, Aesthetics and Usability as stated in Theoretical Framework. The relevant elements then tested either using quantitative or qualitative methods. Based on the Jesse James Gareth's (2003) model of user experience, there are a few processes that have to be conducted starting from strategy, scope, skeleton, structure, to surface. Each process is supporting each other and should be completed one at a time. Based on the questionnaire in Pre Test session, 62.5% from 32 respondents stated that Design 2 would provide them an engaging experience. 85% of them decided that they will visit Pusat Sains Negara (PSN) after browsing the website. In conclusion, an engaging website experience will influence visitors to actual physical visits thus increasing the credibility of the institution. In order to adapt exceptional experience in science museum website, it is always a good practice to consider the latest trends and new approaches in web design. For this research, both innovation scrolling and parallax effect has been integrated into the website design. As a results, the website manage to offer a new dimension in terms of engaging user experience to visitors.

7 Conclusion

Experiments uncover that the combination of design principles, aesthetic, and usability has the potential to improve the quality of science museum websites. The success of a website is not only judged by aesthetic value but also on its ability to provide an engaging experience to users. In this paper, researchers presented a framework on how to provide an engaging website experience for Science Museum which consists of data from design principles, usability and aesthetics evaluation. Web designers should experiment with lots of new ideas to offer something unique. Continuous research on how to provide best user experience should be encouraged. Hopefully this research will spark more interest from designers or developers to join in the bandwagon in terms of producing a more engaging websites design and development.

References

1. Alben, L.: Quality of experience: defining the criteria for effective interaction design (1996)
2. Barnett, L.A.: Playfulness: Definition, Design, and Measurement. Play and Culture 3 (1990)
3. Blythe, M.A., Overbeeke, K., Monk, A.F., Wright, P.C.: Funology: From Usability to Enjoyment (2003)
4. Collis, B.: Information Technologies for Education and Training (2002)
5. Fogg, B.J.: Stanford Guidelines for Web Credibility: Persuasive Technology Lab. Stanford University (2009)
6. Adelsberger, H., et al. (eds.): Handbook on Information Technologies for Education and Training (2005)
7. Jacobsen, T.: Individual and group modelling of aesthetic judgment strategies. British Journal of Psychology (2004)
8. Garrett, J.J.: The Elements of User Experience: User-Centered Design for the Web (2002)

9. Graham, L.: Basics of Design: Layout and Typography for Beginners, 2nd edn. Delmar (2005)
10. Nielsen, J., Tahir, M.: Homepage Usability: 50 Websites Deconstructed. New Riders, Indianapolis (2001)
11. Pandir, M., Knight, J.: Homepage aesthetics: the search for preference factors and the challenges of subjectivity. Interacting with Computers (2006)
12. Park, S., Choi, D., Kim, J.: Critical factors for the aesthetic fidelity of web pages: empirical studies with professional web designers and users. Interacting with Computers (2004)
13. Chen, Q., Wells, W.D.: Attitude toward the site. Journal of Advertising Research (1999)
14. Fazio, R.H., Chen, J., McDonel, E.C., Sherman, S.J.: Attitude accessibility, attitude-behavior consistency, and the strength of the object-evaluation association. Journal of Experimental Social Psychology (1982)
15. Shenkman, B.O., Jonsson, F.: Aesthetics and preferences of web pages. Behaviour & Information Technology (2000)
16. Krugg, S.: Don't Make Me Think! A Common Approach to Web Usability, 2nd edn. Paperback (2006)
17. Ma, Q., Liu, L.: The Technology Acceptance Model: a metaanalysis of empirical findings. Journal of Organizational and End User Computing (2004)
18. Tilghman, B.R.: Reflections on aesthetic judgement. British Journal of Aesthetics (2004)

Lessons Learned from Designing Non-traditional Interfaces for Educational Applications in South Africa

Michael Wolf

Formula D interactive, 71 Harrington Street, Cape Town, South Africa
michael@formula-d.com

Abstract. As a specialised design consultancy for interactive learning environments and tools, Formula D interactive has gained valuable project experience in designing nontraditional interfaces for digital educational content and tools in the culturally diverse context of South Africa. The aim of this paper is to share the company's experience in the field using prominent examples of their recent work, related research and user testing in order to discuss the merit of large-scale interactive surfaces, gesture-based and tangible interfaces in culturally diverse contexts. The company's work includes interactive displays for science centres and museums as well as digital learning tools for classroom environments.

Keywords: Non-traditional interfaces, interaction design, tangible interaction, gesture-based interaction, Locomotion interfaces, interactive surfaces, multi-touch, interactive learning environments, HCI examples.

1 Introduction

The praxis and scope of designing interactive learning environments goes beyond online platforms for learning content or interactive whiteboards. Specifically nontraditional interfaces [2] create opportunities for computer supported co-located learning and accessibility of technology tools by audiences with little experience with traditional computer interfaces. In the following project examples of Cape Town based design firm Formula D interactive (www.formula-d.com). the author highlights the key considerations in the design process, challenges and lessons learned when designing nontraditional interfaces for learning applications in South Africa.

2 Examples of Designing and Testing Nontraditional Interfaces and Lessons Learned

2.1 Multitouch Surface at the Two Oceans Aquarium in Cape Town

In 2008 our company designed a 100" rear-projected multitouch wall for the Two Oceans Aquarium in Cape Town (Fig 1). The goal of the exhibit was to create an immersive interactive landscape to encourage visitors of the aquarium to enquire and learn about habitats, calls and threats of local frog species.

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 755–763, 2013.

Fig. 1. Frog multitouch wall, **Fig 2.** Frog wall GUI

Users find themselves looking at a photorealistic animated landscape. By touching various hotspots they activate one of the habitats, revealing a 360 Deg panoramic view of a typical area. Now, the challenge is to discover the frog species living in the environment. When a frog icon has been found and touched, a window with description and images, video clips and the call of the frog appears. Screen objects can be moved around by dragging and dropping them, for example, to individual eye-level. Some objects can also be scaled using two fingers or both hands. As multiple frog sounds can be simultaneously activated and deactivated, visitors may collaborate orchestrating “frog symphonies”, broadcast via a directional overhead speaker.

The frog wall sported various novel features which had not been seen in South Africa at the time and thus presented itself as a good case study for nontraditional HCI as it was exposed to some thousand users from various cultural backgrounds every day at the Two Oceans Aquarium.

Lessons Learned

- At first, many people of all audiences did not identify the large screen as an interactive surface. In addition to on screen instructions, additional instructional signage was required even though using bold icons as touch points in the GUI (Fig 2). Touch screens were quite common at the time in South Africa (for example in ATM’s which had been in use across all levels of the working society), so we concluded that various contextual elements did not support the user interaction although the GUI was carefully crafted:

  The screen size did not correspond with the conceptual model people had of a touch surface at the time.
  - Apart from a photo booth, the screen was the only interactive exhibit in the aquarium.
  - The first impression of the exhibit design was reminiscent of the live exhibits at the aquarium (fish tanks). These exhibits are usually not being touched or touching them is prohibited.

- The majority of people in South Africa at the time were unfamiliar with the concept of multitouch (the iPhone had not yet been released in South Africa), so these features were hardly used. The two finger scaling gesture, although indicated by GUI elements (Fig 2), was only slowly adopted by the visitors in 2009, when this interaction entered mainstream through iPhone and iPad. Less privileged communities without access to these devices did not access this feature at all unless they were instructed by staff or observed other visitors.
- Users also mostly did not understand that the content windows could be moved around although this feature is very common with desktop computers.
- In turn, once the screen was in use by at least one user, other visitors frequently joined in. The size of the display seemed to suggest that multiple users are possible.

In summary, our observations seem to confirm that the context and size of the display was in conflict with the conceptual model of touch screens and desktop computers, which were familiar to at least a part of the audience. This resulted in certain GUI elements common to desktop applications not being identified. On the other hand, this contextual distance from traditional interfaces could have led to users embracing the multi-user functionality almost naturally.

2.2 Locomotion Interface at Cape Town Tourism

In 2009 we installed our first 6 metre by 2 metre interactive wall at the Cape Town Tourism information centre, which uses visitor's body motions as a means to navigate digital content. The wall projection showcases 40 activities and sights in the Cape region, set against an animated backdrop of interchanging panoramic representations of iconic areas around the Cape. When visitors step in front of the wall, and align themselves with one of the projected icons on the wall, they trigger animations, sounds, and pop-ups with information on the selected attraction (Fig. 3, 4).

Fig. 3,4. Interactive wall for Cape Town Tourism Visitor Centre

Limitations of the Interaction. At the time we used infrared cameras for blob detection not only for multitouch displays, but also for interactive floor and wall projections (Microsoft Kinect had not been released yet). The overhead 2D-tracking of user activity on the floor only provided x and y positions as well as blob size and grouping as a means of user input. Since the floor in front of the wall is also a passage area in the centre we had to differentiate between intended and accidental users. In order to avoid accidental inputs, we decided to limit the interaction area to a designated floor strip in front of the wall and only worked with object movement on the x-axis, which further limited the user control in the application. We used a vertical colour bar which moved with the user across the wall to make the user aware they are influencing the environment. Next, we decided to trigger content when a user stopped at a certain area as this may suggest an interest in the special area.

Lessons Learned

- Digital “real estate”: There are many challenges and limitations when multiple users interact in a shared digital environment which is based on one shared output. Online, or in AR scenarios where users have individual devices or headsets, a subjective render of the environment with personalised content is possible. However, if there is only one shared output, the “real estate” of the digital space needs to be shared just like our physical spaces.
- In the design process of this project we had a lot of debate around managing the physical interaction area and its corresponding digital space for an unknown number of simultaneous users. We were concerned about conflicts arising between users who aimed at triggering events in proximity to each other. This resulted in a quite rigid partitioning of the interface area with finite positions of trigger points, so we could be sure there would not be spatial conflicts between interacting users.
- However, this concern has proven to be unjustified since the natural behaviour of people in communal space is to keep a minimum distance between each other unless two people are grouped, like a couple walking arm in arm or an adult with a child. It needs to be noted that this minimum distance between people in physical space differs in different cultural environments. However, we observed that a “spatial etiquette” is maintained when multiple users interact in large interactive environments such as the interactive tourism wall or the Frog Touch wall. People do generally keep distance physically and virtually.

2.3 Gesture Controlled “Point Screen” for the Centre for Public Service Innovation

As part of a larger installation of various interactive displays for the Centre for Public Service Innovation (CPSI) near Pretoria, we created several user interfaces for a hand and finger-tracking device developed by HHI Fraunhofer in Berlin (Fig 5). iPoint accurately tracks one or multiple finger gestures, which can then be calibrated to navigate a Graphical User Interface. We developed two applications for this system for an educational exhibition environment, one information kiosk with best practise examples of innovation in public service, one assessment of the users’ understanding of innovation principles.

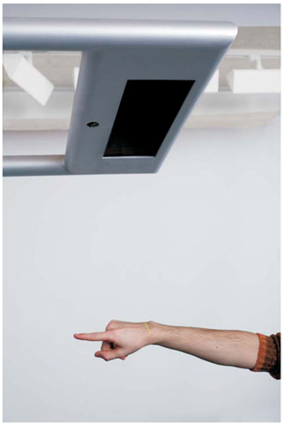


Fig. 5. iPoint tracker by HHI Fraunhofer, **Fig 6.** Problems with the gesture application

Lessons Learned

- Challenges with the conceptual model of the interaction: Although pointing surely is one of the most archaic gestures as it can already be observed in small infants, users presented with a screen and the invitation to point at it behave all but "natural". Even when instructed to point at the screen "naturally" many users bent their arms in anticipation of triggering a sensor in the black box above or underneath them.
- It seems "unnatural" to users to interact with a screen by pointing at it. Thus, the gesture itself needed to be instructed and could only be applied after reflection in the new context (Fig. 6).
- The trigger problem: Most computer interfaces separate pointer and input action. This gives users a chance to reflect on their choice before executing an action. A gesture based point interface does not have a mouse click, so an alternative way of selecting an action had to be invented. Already 10 years ago, I had worked on pointing devices using electric field sensing (EFS) technology at MARS Lab (Fraunhofer Institute for Media Communication) in Germany [3]. Various ideas and concepts were thought out at the time, such as multimodal interface solutions which combined finger pointing action with finger snapping or voice commands. However, the added complexity of the interaction always felt unsatisfactory since it undermined the simplicity of the pointing gesture.

  Thus, for our South African client, we looked at GUI solution that did away with clicking and worked solely with roll-over like the website experiment *Don't click it* [1].

Our first iPoint GUI (Fig. 7) was a structure of unfolding and collapsing content display areas which open and close according to the position of the cursor. However, what seemed a viable solutions for a Desktop computer interface with a mouse pointer, did not work as a finger-pointing device. After selecting content, users had to leave their arm in the same position as long they needed to absorb the content, which caused great discomfort for our users. A subsequent version was designed in a way that users could lower their arm without triggering other content areas. But even this solution resulted in many accidental selections. The current design comprises of a time-delayed activation, a visual countdown, which activates a link only if a user remains in one position for a few seconds. A similar solution is featured in Microsoft Kinect games.

Fig. 7. iPoint “Don’t click” GUI, **Fig 8.** Sample of final GUI solution

- Pointing Ergonomics

Even with the time-delayed activation functionality, we still faced ergonomic problems. It was uncomfortable and tiring to navigate the interface even if just for a few minutes at a time. We realised that we had spent a lot of time thinking about trigger mechanisms, but had not yet looked closely at the specific ergonomics of pointing with hand and arm. After a few tests we found that although user managed to navigate traditional GUI’s with a horizontal menu structure (Fig 9), keeping arm and hand on a small target for the necessary 2 seconds until the link activated added a fair amount of discomfort. The second generation of tests was conducted with a checkerboard interface (Fig 10), which not only offered larger hit areas, but also used the entire screen space as the interactive area. This made the interaction more fluid and users felt more in control. However, at the same time, the absence of a resting area pressured users to make a quick selection, which made the interaction uneasy. The current interface and best solution to date is a circular GUI (Fig 8, Fig 11), which follows the natural ergonomics for pointing with the hand, since it supports the circular movement of the wrist. The movements feel natural. At the same time, movements can be reduced to an absolute minimum; the interface can be navigated with only the index finger moving. With regard to the trigger area, we found that adding large interactive areas and rollover states added to the experience of control and ease, whilst the trigger areas or buttons needed to be separated from the active areas, so the trigger timer would only be activated if users intended to do so.

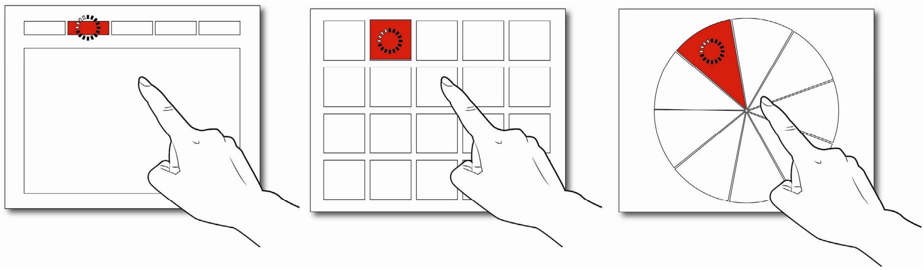
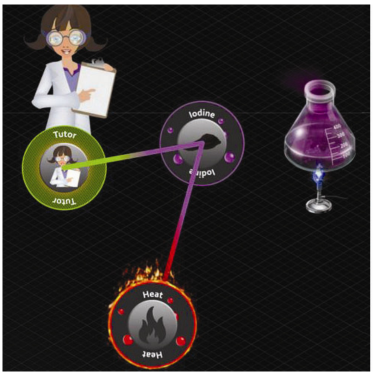


Fig. 9., Fig 10., Fig 11. Evolution of the point GUI

2.4 The Virtual Chemistry Lab Table (VCLT)

A few years ago Formula D interactive started the development of the Virtual Chemistry Lab Table (VCLT), a tool to help learners understand the basics of Chemistry through hands-on interaction. Once connected to a standard computer and screen, and the software is installed, the Virtual Chemistry Lab Desk allows learners to arrange physical objects on a surface in order control the software (Fig 12). Learners then explore the digitally simulated experiments using an array of tools similar to the ones in a real chemistry lab by adding other objects in proximity to each other. A simple content management system expands the functionality of the lab from a simulator to a documentation and presentation tool. Here, learners embed their own knowledge or test results (from real lab experience or secondary research) and embed content, such as images, video or text within the application. Now, they can share and discuss their findings with peers in a classroom setting. The VCLT is built on top of the reactivision platform (http://reactivision.sourceforge.net/) with custom built hardware and content development system.



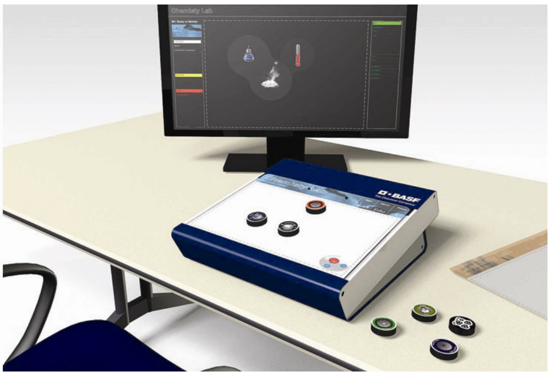


Fig.12. VCLT GUI**, Fig 13.** Desktop console concept render

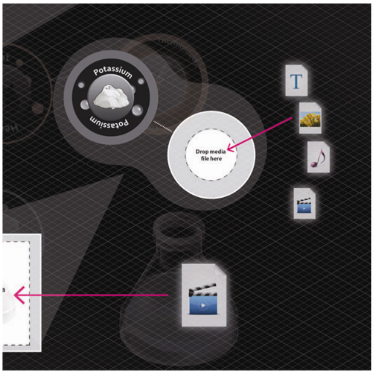


Fig.14. VCLT CMS**, Fig 15.** User testing at Wynberg Girl's Junior School

Lessons Learned

- A virtual chemistry lab is a great application for tangible interaction through reference objects. Users can easily transfer the mental model from a real chemistry lab to the simulated one. The interactions are similar yet simplified.
- Initial considerations for the features of the lab made only provision for various content modules with simulated experiments. Various discussions with teachers and students suggested, however, that the lab needed the possibility for user-generated content through a simple CMS (Fig 14). At this point the lab would also become a documentation centre and database, which could eventually be shared with other VCLT's online.
- The VCLT was originally designed as a Science Centre exhibit. The design included a much larger table with a rear-projected screen of 50". The content animations projected around the objects which were laid on top of the glass surface.
  Since many South African schools wouldn't be able to afford a large table, we developed a low cost version which was able to plug into existing hardware and was more suited to a classical classroom scenario.We then experimented with a prototype which separated input and output (Fig 13). We expected problems with the interaction and were surprised when young learners of different cultural backgrounds acted very confidently when operating the nontraditional input device in combination with a traditional output device (screen).
- For our initial user testing (Fig 15) we selected a girl's junior school in Wynberg, a suburb of Cape Town. The pupils at the school come from culturally diverse backgrounds. We observed 8 girls in groups of two for 15 minutes each, and a group session with teacher. Approximately 50% of the testing group has access to a computer at home.
  The observations confirmed that:
  - The interface was easily understood without instructions by the vast majority of the users including users who have no regular access to computers.
  - Only one girl had to be informed of the possibility of combining different objects, after she had placed single objects onto the lab one by one.

- Most users were able to memorise the experiments they tried out.
- Observations of user interaction confirmed that the VCLT invites collaboration since the tangible interface objects can easily be shared and jointly operated.

In various interviews the users commented positively on the tool:

- They liked that they could "see what they are doing" through the strong representation of the current active state in the interface (objects/ingredients on the table).
- They highlighted that they enjoyed the hands-on approach, as opposed to just listening to the teacher.

3 Conclusion

A recurring dilemma in designing nontraditional interfaces is that the designer's intention of supporting more meaningful and user-friendly interaction through more "natural" interaction frameworks (like gesture, locomotion or tangible interfaces) is prone to failure when the user's conceptual model of how technology works is predominantly based on traditional interfaces they have been exposed to.

These preconceptions can only be influenced if designers use strategies that make provision for the context of the user and the environment in which the tool is deployed. Through project work on the Virtual Chemistry Lab Table it became apparent that using a combination of traditional and non-traditional interface elements can be a good strategy to offer the user a familiar environment on one side, building up the confidence needed to engage with new forms of interaction on the other. Beyond these design considerations and strategies, Formula D interactive's project experiences indicate that non-traditional interfaces such as the Virtual Chemistry Lab Table improve co-located learning and collaboration in the classroom and make interactive digital technology more accessible to audiences with no or little experience with traditional interfaces.

References

1. Don't Click It!, `http://www.dontclick.it`
2. Kortum, P. (ed.): HCI beyond the GUI – Design for Haptic, Speech, Olfactory and Other Nontraditional Interfaces. Elsevier Inc., Burlington (2008)
3. Strauss, W., Fleischmann, M., et al.: Information Jukebox – A semi-public device for presenting multimedia information content. Personal and Ubiquitous Computing 7(3-4), 217–220 (2003)

The Bigger Picture: The Use of Mobile Photos in Shopping

Maryam Tohidi and Andrew Warr

Google, Inc.
1600 Amphitheatre Parkway, Mountain View, CA, 94043
{mtohidi,andywarr}@google.com

Abstract. Mobile phones are becoming, if not already, an integral part of our lives. They have a wide range of applications, such as communication, gaming and commerce. Shopping in particular is a rapidly growing domain. Today, shoppers use their phones to make more informed shopping decisions by researching products and merchants, save money using price comparison, mobile coupons and daily deal apps, even purchase products directly on a mobile device. While mobile commerce and shopping apps are in the spotlight, one area that has received little attention is the role of the native capabilities of a mobile phone, such as the mobile camera, in the shopping process. This paper demonstrates the key role mobile photos play in the shopping process, documenting use cases, practices and pain points, and informing opportunity areas for mobile shopping applications and services.

Keywords: Mobile, Phones, Cameras, Shopping, Photos.

1 Introduction

In recent years there has been a steady rise in the use of mobile phones. At the end of 2012 an estimated one billion smartphones were in use worldwide [1]. The adoption of mobile devices and their affordances such as advanced computing capabilities and connectivity have enabled new areas of application; one such area is mobile-supported shopping.

Academic research on mobile shopping has primarily focused on mCommerce --- shopping for and purchasing items using a mobile device [7]. Hillman et al. [3] studies the everyday routines of regular mobile device shoppers and identifies a large variety of mobile shopping activities, such as looking for products, comparing prices and purchasing items. Whilst such research has deepened our understanding of mCommerce, little in-depth research has considered the role of mobile phones when shopping in the real world.

According to US census data, eCommerce - a superset of mCommerce - accounted for only 5.5% of all retail sales in the US in 2013 [8]. With a large majority of commerce still taking place offline, it is important to understand the role of mobile

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 764–771, 2013.

devices in supporting shopping in physical stores. Pew Research reported that 6 in 10 mobile phone owners used their phone inside a physical store for assistance or guidance on a purchasing decision [6]. Nielsen and Smith [5, 6] demonstrate the prominence of several common activities such as calling a friend or family member for advice, looking up reviews, looking up price, locating stores and researching products, to name the most common activities. However, some key use cases for mobile phone usage in-stores were omitted in these surveys, resulting in a lack of awareness of their importance.

One area that has received little attention in the mobile shopping literature is the use of mobile phone photos to support shopping. Häkkilä et al. [2] have examined the use of camera phones in general, identifying that the most common functional (i.e. task driven) use for mobile phone photos was to support shopping activities.

In this paper, we explore the application of mobile phone photos in supporting shopping activities and present findings from a quantitative survey as well as a qualitative analysis of photos, identifying motivations, use cases and pain points.

2 Quantitative Research

To examine the relative importance of shopping-related mobile phone photos we conducted two surveys in November 2012 and February 2013.

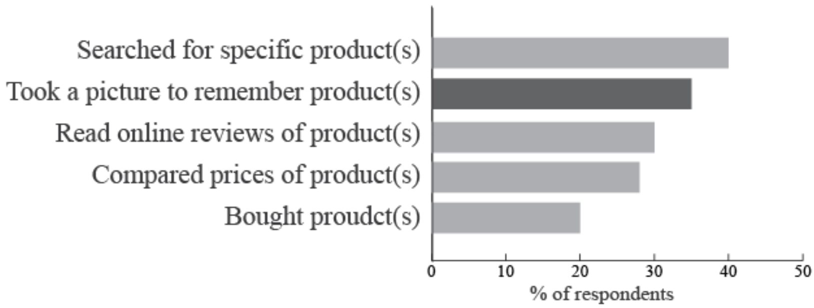


Fig. 1. Survey 1, top mCommerce activities

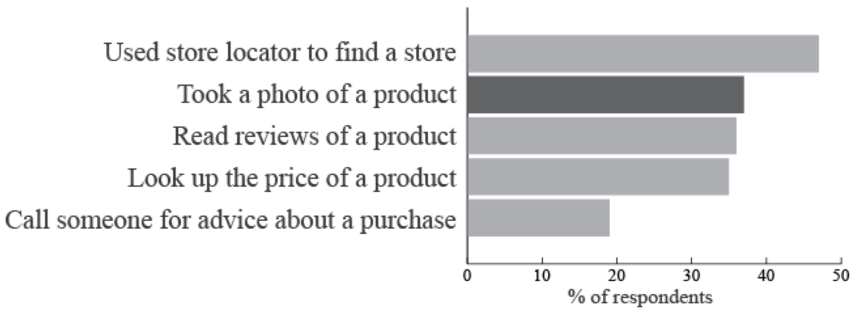


Fig. 2. Survey 2, top in-store mobile shopping activates

In the first survey, 2005 respondents, representative of the US population of online users, answered the multiple choice question 'Which of the following shopping activities did you do on your mobile phone in the past 2 weeks?' The choices included the most frequent *mCommerce* activities identified in previous research [3, 5], as well as 'took a picture to remember a product'. From those who had done any shopping activities on their phones (503), taking a picture to remember a product ranked among the top.

In the second survey, 1339 respondents, representative of the US population of online users, answered a similar question; this time the choices included the most frequent *in-stores mobile shopping activities* identified in previous research [5, 6], as well as 'took photo of a product'. From those who had done any shopping activities on their phones (360), taking a photo of a product ranked among the top again.

Our surveys showed that taking shopping-related mobile photos is comparable to mobile shopping activities previously identified as most frequent.

3 A Study of Shopping Photos

To understand how mobile photos are used in shopping we collected real examples of shopping-related photos. We posted requests on social and miscellaneous mailing lists at a large US technology company, asking recipients to browse their mobile photo gallery for photos taken related to a shopping activity in the last month, and to email us such photos accompanied by a caption stating why they took the photo.

We received a total 285 photos and associated captions from 91 respondents (58% female). Their ages ranged from under 20 to over 50, with a median age range of 30-39 years. Respondents were from across the US, representing a range of jobs, both technical and non-technical.

Furthermore, we asked the participants to elaborate on their practices and pain points around taking these photos in a follow up survey. 80 of the 91 responded to this survey.

3.1 General Observations

As suggested by previous research [2], many of the photos in our study were taken as a memory-aid. Since their mobile phones are almost always with them, participants used their phone as a repository of to-dos, useful references, reminders and evidence related to their shopping. One participant said *"I usually forget the grocery list but not my phone. So I just take a photo of it."* Another participant took a photo of a receipt, *"so that I had the receipt saved in case I lost it."* Since their primary purpose is functional [2, 4], little attention is given to the composition, lighting and other artistic values of these photos.

We also saw evidence of 'bursts' [2] of photos, taken of similar products, related to the same shopping purpose. For example, one participant had taken 8 photos of floor lamps to show her boyfriend and another participant had taken 7 photos of TV options he had seen at Costco to research later. In our follow up survey 61% of respondents confirmed that they have taken at least one burst of 3+ photos related to the same shopping intention.

3.2 Photo Analysis

We took a bottom-up approach to analyzing the photos based on what was pictured (i.e. content) and why they were taken (i.e. intent). Two researchers iteratively developed a coding scheme covering these dimensions. Before coding the entire collection, a random sample of 10% of the photo/caption pairs was coded by two independent judges. The coding scheme had an inter-judge reliability of 80%. Table 1 presents these codes.

Table 1. Coding scheme, frequencies and % of all photos

Content codes	# (%)	Intent codes	# (%)
Full product	154 (54%)	To remember	146 (51%)
Label	121 (42%)	To share	124 (44%)
Price	68 (24%)	To consider	102 (36%)
Something owned	45 (16%)	To buy	96 (34%)
Barcode	40 (14%)	To inspire	44 (15%)
Partial products	35 (12%)		
Self	10 (4%)		
Shopping list	9 (3%)		
Receipt	7 (2%)		
Advertisement	7 (2%)		

3.3 Content

Physical Products: Taking a photo of the actual product was most common, with 54% of photos showing the full product (Figure 3a), and an additional 12% showing a partial product (Figure 3b) - often a close-up focusing on a specific aspect of the product, such as the texture of a scarf. Furthermore, taking a photo of a product already owned accounted for 16% of photos. (Figure 3c)

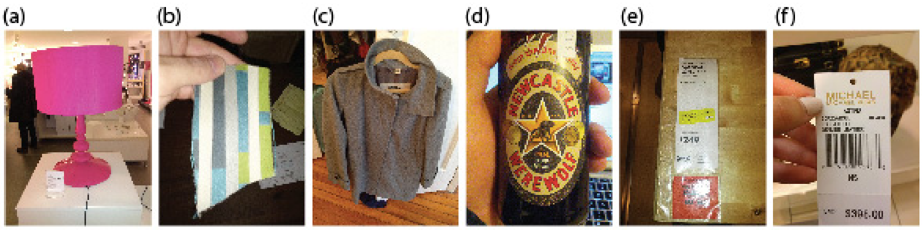


Fig. 3. Examples of shopping photos

Information about Products: Labels were the 2nd most common content type, accounting for 42% of photos (Figure 3d). Labels included textual or numeric information describing the product (e.g. wine bottle labels, clothing labels and nutritional information). Price and barcodes were also quite commonly pictured, appearing in 24% and 14% of photos respectively (Figure 3e and 3f).

We also saw photos of oneself (often trying on, or setting context for a product seen in a physical store), shopping lists, receipts and advertisements.

3.4 Intent

The accompanying caption described the intention behind taking a photo. Analyzing these captions, allowed us to categorize the primary motivations (Table 1). Shopping photos were taken as a means to remember (51%), share (44%), consider (36%), or buy a product (34%), or to serve as an inspiration for a future purchase (15%).

These intentions often overlapped (shown in Table 2), for example, from the photos taken with the intention 'To Remember', 49% were also intended 'To Buy', and '29% 'To consider' a product.

Table 2. Overlap of intentions, highlighting the most common

	To share	To consider	To buy	To inspire
To remember	10%	**29%**	**49%**	9%
To share		**52%**	9%	**22%**
To consider			5%	1%
To buy				**14%**

4 Discussions

Investigating the intent behind the photos allowed us to drive the common use cases in which mobile photos support shopping.

Just over half (51%) the photos were taken with the intention to remember something, helping with the *organizational aspects of shopping*.

Shopping List: 49% of photos taken with an intent 'To remember', were also intended 'To buy' a product. These photos served as visual shopping list.

Shortlist: 29% of photos taken with an intent 'To remember', were also intended 'To consider' a product. These photos, often taken of similar products, helped keep a shortlist of products under consideration, to research or compare later.

Archive: Another important use case of photos taken 'To remember' was archiving evidence or important references, such as receipts and shipping labels.

The *social aspect of shopping* is another important dimension supported by photos. 43% of all photos were taken with the intention to share with someone. These photos were often shared through MMS, email, or shown to someone in person.

Second Opinion: The primary reason for sharing a photo was to get feedback on a product under consideration (52%). These photos were often shared with a spouse, a close friend or family member, asking for a yay or nay response. There was only one example of sharing with a group to get collective feedback for a mutual purchase.

Hint, Hint: 22% of photos taken with an intent 'To share', were also intended 'To inspire' someone else to buy something. For example, one participant sent a photo of her wedding band to her husband as a reminder for her anniversary gift.

Saw this, thought of you: Another reason for sharing photos was to maintain or enhance personal relationships. For example one participant shared a photo of a beer bottle label featuring a werewolf with a group of friends playing a computer game fighting werewolves. Humor also played a role in this type of sharing. One participant took photos of owl figurines to send to a friend who is scared of owls. Another posed for a photo with an infant outfit held in front of her to share with her family for a laugh. This was the only category of sharing, where group sharing was very common.

Other sharing scenarios included expressing frustration with a bad shopping experience, providing proof of a purchase, or setting context for product recommendations.

4.1 Mapping of Content and Intent

A mapping of photo intent against photo content illustrates some of the reasons why people took photos of the different aspects of a product. In Table 3, each cell indicates the percentage of photos taken of a content type (row), with a given shopping intent (column). Note than since 'To remember' almost always went hand-in-hand with another intention that was more specific to shopping, we have simplified this graph by focusing on the other 4 intentions.

Table 3. Mapping photo intent against content – highlighting the most common

	To share	To consider	To buy	To inspire
Full product	**62%**	**46%**	21%	16%
Label	30%	31%	**43%**	15%
Price	34%	**49%**	29%	9%
Something owned	24%	0%	**78%**	16%
Barcode	5%	38%	**58%**	10%
Partial product	29%	17%	**49%**	9%

4.2 Inferring Intent from Content

Table 3 highlights usage patterns that could inform our ability to infer shopping intent from the photo content alone. As show above, when the photo shows a full product, the most common shopping intentions are to share the photo with someone else (62%), or to consider the product further (46%). Photos focused on partial products however were more often tied to buying (49% compared to 21% for full product).

Taking a shopping-related photo of something they already owned, was very often tied to a buying intent (78%) in order to replace or complement their existing product.

Similarly, when the primary content of the photo was a product label or barcode, there was often a buying intent (43% and 58% respectively). Price tags on the other hand, tended to indicate a product under consideration (49%).

4.3 From Inspiration to Consideration to Purchase

The shopping stories (captions) shared with us often spoke of longer term purchases, where the photos connected the dots between different phases of shopping. *"I've been shopping online for boots for a couple of months, and I'd been showing my husband a lot of pictures. He sent me several [mobile] photos of boots he saw on two shopping trips, while he was out shopping for something else. We followed up by text/phone. We ended up buying the red boots."*

Inspiration-related photos were often taken of objects seen in the real world (e.g. *"a lamp I liked in a restaurant"*), ads or magazines (e.g. *"lighting feature in a design magazine I want to DIY"*), and collections of products (e.g. *"a nursery room in Pottery Barnes"*). 61% of inspiration-related photos were to be shared with someone else, either to suggest gift ideas (e.g. *"he might like this for Christmas"*), or as a self-appointed personal shopper, curating based on the recipient's interests or needs (e.g. *"Cadbury mini-eggs are in stores already! I know you love them"*).

Consideration-related photos were often taken in 'bursts' to record references to products that the shopper intends to research or compare. Furthermore, consideration and sharing often went hand in hand. For example, a group of bridesmaids took several photos of dresses to send to the bride for her consideration. Apparel and furniture were the 2 most common product categories that participants shared with others to get a second opinion.

Purchase-related photos were centered around specific product(s) to buy later, either online or in-store. About half of these photos were taken in a store to buy online later, *"Liked this area rug at Home Depot but in-store inventory was frayed. I went home and ordered it online"*. In many cases there was a hope to find the product at a cheaper price, *"Ski Jacket that was on sale at REI - I wanted to see if I could find a better price online."* There were also photos to help future in-store purchases, such as photos of shopping lists and recipes.

4.4 Pain Points

Although shopping-related photos are often taken as a memory aid for future actions, it is not uncommon for the photos themselves to be forgotten. *"I took these pictures to see whether that item has good reviews. Then forgot about it and took no action."* In the follow-up survey, 83% said they have taken at least one photo with the intention to refer to it later, but they forgot to do so. Shopping-related mobile photos tend to stay in the mobile gallery, among other photos, making them harder to remember, find and organize over time. (72% said they don't go back to delete the shopping related photos from their mobile gallery.)

When asked what the biggest challenges regarding these photos were, the primary pain points identified in the follow up survey were remembering to go back to the photos (59% of respondents), finding a photo when you need it (53%), finding the product at a later date (33%).

5 Areas of Opportunity

Our study suggests that, while there is plenty of usage and applications for mobile photos in the shopping process, there is room for better connecting the dots between the two mediums. The biggest challenge for users is that they forget the photos, or have a hard time finding them when needed. There seems to be an opportunity area for tools to help identify, organize and surface these photos at the right times and places so that they can better support shopping.

6 Conclusions

This paper highlights the importance of mobile photos in supporting shopping activities, an area previously overlooked in related research. We have proposed a taxonomy for the content and intent of shopping-related photos, identified key use cases, common pain points and suggested an opportunity area for further development.

References

1. Bicheno, S.: Global Smartphone Installed Base Forecast by Operating System for 88 Countries: 2007 to 2017, `http://www.strategyanalytics.com/default.aspx?mod=reportabstractviewer&a0=7834` (accessed June 2013)
2. Häkkilä, J., Huhtala, J., Sarjanoja, A., Schmidt, A.: Price tags, maps, recipes: mobile phone photos for functional purposes. In: Proc. of NordiCHI 2012, pp. 41–44. ACM, New York (2012)
3. Hillman, S., Neustaedter, C., Bowes, J., Antle, A.: Soft trust and mCommerce shopping behaviours. In: Proc. of MobileHCI 2012, pp. 113–122. ACM, New York (2012)
4. Kindberg, T., Spasojevic, M., Fleck, R., Sellen, A.: The Ubiquitous Camera: An In-Depth Study of Camera Phone Use. IEEE Pervasive Computing 4(2), 42–50 (2005)
5. NielsenWire, Mobile Devices Empower Today's Shoppers In-Store and Online, `http://www.nielsen.com/us/en/newswire/2012/mobile-devices-empower-todays-shoppers-in-store-and-online.html` (accessed June 2013)
6. Smith, A.: In-store Mobile Commerce During the 2012 Holiday Shopping Season, `http://pewinternet.org/Reports/2013/in-store-mobile-commerce.aspx` (accessed June 2013)
7. Stafford, T., Gillenson, M.: Mobile Commerce: What it is and What it Could be. Communicaitons of the ACM 46(12) (2003)
8. U.S. Census Bureau Report, Quarterly Retail E-Commerce Sales, `http://www.census.gov/retail/mrts/www/data/pdf/ec_current.pdf` (accessed June 2013)

Human-Centred Design: What's Happening in African Industry?

Susan Dray[1], Phil Barrett[2], Mark Kamau[3], Edna Chelule[4], and Tamara Alsheikh[5]

[1] Dray & Associates, Oakland, USA
susan.dray@dray.com
[2] Flow Interactive, Cape Town, South Africa
Phil@userexperience.co.za
[3] iHub UX Lab, Nairobi, Kenya
Mark@ihub.co.ke
[4] DStv Online, Johannesburg, South Africa
Edna.Chelule@dstvo.com
[5] Google EM, Zurich, Switzerland
t.alsheikh@acm.org

Abstract. There has been much written about the current 'state of the art' of user experience (UX) and interaction design in the both the academic and the business press in North America and Europe. But where is the African voice? What is happening in these fields in Africa? This panel is intended to explore the African playing field of UX and interaction design. The panellists are all well-known and well-respected within the field and within Africa. We plan to discuss what we see as the current 'state of the art' as well as the challenges, especially the unique challenges, and opportunities that we face in advancing this profession within Africa.

Keywords: User experience, interaction design, human-centred design, Africa.

Panel on Community Collaboration

Anicia Peters[1,3], Darelle van Greunen[2], Gereon Koch Kapuire[3], Anupama Saxena[4], Dan Orwa[5], Bongiwe Dlutu-Siya[6], and Nicola J. Bidwell[7]

[1] Iowa State University, Ames, USA
[2] Nelson Mandela Metropolitan University, Port Elizabeth, SA
[3] Polytechnic of Namibia, Windhoek, Namibia
[4] Guru Ghasidas University, Bilaspur, Chhattisgarh, India
[5] University of Nairobi, Nairobi, Kenya
[6] Rhodes University/Mankosi Community, Grahamstown, SA
[7] University of Pretoria, Pretoria, SA

anpeters@iastate.edu, Darelle.vanGreunen@nmmu.ac.za, gkapuire@polytechnic.edu.na, anupama66@rediffmail.com, dorwa@uonbi.ac.ke, {bongiwe.dlutu,nic.bidwell}@gmail.com

Abstract. Working with local communities and gaining an intimate knowledge of a community life can make significantly improve the efficacy and appropriateness in the research and design of interactive systems. However, it is not always easy to substantially engage a community in research and design, due to constraints and barriers both within and without the project's domain. Common constraints and barriers include navigating cultural, socio-economic and political issues, power relations and trust; poor or unfamiliar technical infrastructure; time and funding constraints; and, organizational requirements and expectations of funding agencies and different project partners.

Five panelists will share their experiences in engaging underserved communities in Africa and India in community-based participatory designs and research. Their discussion seeks to enrich our understanding and practices in collaboration by addressing the following themes:

Effective approaches to and guidelines for collaborating with underserved communities

Strategies to overcome barriers in community-based co-design and co-research

Stakeholder roles

Ways that theoretical concepts in African philosophy can assist in configuring collaboration

Issues raised in a workshop on community participation, to be held at INTERACT 2013 preceding the panel

Panelists will be: Anupama Saxena, Gereon Kapuire, Bongiwe Dlutu-Siya, Dan Orwa Ochieng, Nicola J. Bidwell, Darelle van Greunen (Moderator).

Keywords: community collaboration, HCI4D, human-computer interaction, HCI, development, co-design, participatory design.

Seeing ‘the Obvious’ as Subject and Object of Cross-Cultural HCI Design

Jose Abdelnour-Nocera[1], Heike Winschiers[2], Torkil Clemmensen[3], and Christian Sturm[4]

[1] Sociotechnical Centre for Internationalisation and User Experience, Uni. of West London, UK
[2] School of Information Technology, Polytechnic of Namibia
[3] Copenhagen Business School, Denmark
[4] Hamm-Lippstadt University of Applied Sciences, Germany
jose.abdelnour-nocera@uwl.ac.uk, hwinschiers@polytechnic.edu.na, tc.itm@cbs.dk, c.sturm@arolis.com

Abstract. The aim of this panel is to encourage debate on the idea that the real challenge for cross-cultural design in HCI is part of a science aimed at making the ‘obvious’ visible. This idea will be presented to the audience from four different perspectives. The different groups involved in the design endeavor operate through their own common sense realms, which tend not to be shared across their boundaries. In cross-cultural design settings, key requirements for the things we design and key assumptions about how we design them tend to be hidden in realities for which we have limited access. This is not only a problem of physical access but also of symbolic access to local and professional cultures. By looking at their language, we can see how other stakeholder groups view their world. What we cannot see easily is their ‘point of view’, i.e. the frames full of values, tacit knowledge and expectations that shape language and action.

The panelists will talk about the boundaries of concepts and tools in HCI and Cultural Theory used in cross-cultural design and their effectiveness in unpacking ‘the obvious’, i.e. enabling access to the different ‘points of view’ in the design process. It is envisaged that the outcome of the panel session will be an increased awareness of the panelists’ proposals to address the problems faced in cross-cultural design. After more than a decade of research into culture and HCI we hope the discussion generated in the panel will reflect a high level of nuance and maturity with clear take-away for researchers and designers.

Keywords: cross-cultural design, HCI, interaction design, cultural theories.

What Can Design Laboratories Do?

Thomas Binder[1], Lone Malmborg[2], Per-Anders Hillgren[3], Jörn Messeter[4], Yanki Lee[5], and Sibukele Gumbo[6]

[1] Royal Academy School of Design, Denmark
[2] IT University of Copenhagen, Denmark
[3] Malmö University, Sweden
[4] Cape Peninsula University of Technology, South Africa
[5] EXHIBIT, United kingdom
[6] Siyakhula Living Lab Management Unit
tbi@kadk.dk, malmborg@itu.dk,
{per-anders.hillgren,jorn.messeter}@mah.se,
yankilee@hotmail.com, s.gumbo@ru.ac.za

Abstract. The distinction between design and use is getting blurred in several ways as products and services increasingly are co-created in use, and use and users proliferate and diversify. This has led to a growing interest in staging what we here will call design laboratories that both involve future users in codesign and incubate emergent everyday practices of design in use. Very little can be said about novel designs or technologies before it is seen what the users make of it. This is particularly the case when considering mobile technologies and social media where products seem to act more as an infrastructure for design in use than as a provision of well defined services. At the same time envisioning such new designs without being in close dialogue with the people who are to appropriate them in new everyday practices may easily lead to disappointing results. This has led to a renewed interest in codesign and user involvement in such formats as living labs and codesign workshops. This panel brings together researchers from some of the environments that most vigorously has pursued this line of inquiry. Some of the panelists have suggested the concept of design laboratories as a platform for open-ended explorations of the co-evolution of artifacts and practice. They see the design laboratory as a confined space of controlled experimentation with what can be collaboratively imagined, scalable through its rehearsals of experimental practices. With this definition of these new collaborative formats the panel asks what it is that constitutes a successful design laboratory and what we can expect design laboratories to accomplish in a world of diversity.

Keywords: codesign, participation, living labs, community, open innovation.

Promoting Human-Computer Interaction Values and Practices in Small and Emerging Economies

David Lamas[1], Panayiotis Zaphiris[2], Georgios Christou[3], and Alfredo Terzoli[4]

[1] Institute of Informatics, Tallinn University
Narva mnt 25, 10134 Tallinn, Estonia
[2] Department of Multimedia and Graphic Arts,
Cyprus University of Technology 84848, Limassol, Cyprus
[3] European University Cyprus
6 Diogenes St., 1516, Nicosia, Cyprus
[4] Department of Computer Science, Rhodes University
Grahamstown, South Africa
david.lamas@tlu.ee, panayotis.zaphiris@cut.ac.cy,
georgios.christou@acm.org, a.terzoli@ru.ac.za

Abstract. This special interest group aims to further the understanding of the challenges related to the design and evaluation of digital interactive artifacts in small and emerging economies. Computing is at one of its most exciting moments, playing an essential role in supporting human activities, facilitated by the growing availability of services, devices and interaction modalities. Moreover, with the evolution from the large-scale computing to the contemporary ubiquitous computing, users were brought from the periphery to the center of an emerging pervasive socio-technical system, which pulled the inherent interaction paradigms through the successive waves of the personal, networked, collaborative, mobile, augmented and virtual reality interaction paradigms.

Notwithstanding, in many small and emerging economies, interactive systems are still being designed and evaluated without fully taking into account our cognitive abilities, giving little or no consideration to the ways we perceive and handle information, go about our work and life, create and maintain social relations, and use our cultural context and relate to our environment, thus failing to realized the technology's potential.

Keywords: Small economies, emerging economies, human-computer interaction, interaction design, usability.

Designing High-Quality Surveys for HCI Research

Hendrik Müller[1] and Aaron Sedley[2]

[1] Google Australia Pty Ltd., 5/48 Pirrama Road, Pyrmont, NSW 2009, Australia
[2] Google Inc., 1600 Amphitheatre Parkway, Mountain View, CA 94043, USA
{hendrikm,asedley}@google.com

Abstract. Survey research is widely used in human-computer interaction (HCI) to measure users' attitudes and collect product feedback. At a glance, survey research seems straightforward: it's easy to write questions, and there are many inexpensive tools to field surveys. However, there is a wide gap between quick-and-dirty surveys and surveys that are properly planned, constructed, and analyzed. This interactive full-day tutorial explains all stages of the survey research life cycle, and examines approaches for surveys to meet specific HCI goals.

The tutorial is divided into two parts: During the first half, we discuss the use of surveys in HCI research, and review survey fundamentals (including survey types and examples, a short history of survey research, survey appropriateness in the context of HCI, an overview of the survey life cycle, research goals & constructs, and population & sampling). In the second half, we will examine elements of high quality questionnaire design, including when to use various question types, questionnaire biases and question types to be avoided, considerations for cross-cultural surveys, visual design guidelines, as well as the remainder of the survey life cycle (including testing and optimizing your survey, implementation considerations for online surveys, maximizing response rates, data analysis fundamentals). Interactive exercises and numerous examples are used throughout to engage the audience with the material and to make it immediately applicable to their work.

The content of this tutorial is relevant to individuals from academia, industry, and government with a common desire to further their knowledge of survey research and its uses in HCI. The audience may include user experience researchers, designers, and developers, as well as product managers, and analysts, among others. The tutorial targets both those wanting a foundation to start using surveys and other feedback-gathering methods, as well as those looking to refine their existing survey research efforts. Attendees will gain an appreciation for the breadth and depth of surveys in HCI, combined with keys to conducting valid, reliable, and impactful survey research for their own purposes.

Keywords: surveys, feedback forms, research methods, HCI.

Building Community Collaboration

Anicia Peters[1,3], Darelle van Greunen[2], Heike Winschiers-Theophilus[3], Shilumbe Chivuno-Kuria[3], Kagonya Awori[4], and Anirudha Joshi[5]

[1] Iowa State University, Ames, USA
[2] Nelson Mandela Metropolitan University, Port Elizabeth, SA
[3] Polytechnic of Namibia, Windhoek, Namibia
[4] iHub Research, Nairobi, Kenya
[5] Industrial Design Centre, IIT Bombay, India

anpeters@iastate.edu, Darelle.vanGreunen@nmmu.ac.za, {hwinschiers,schivuno}@polytechnic.edu.na, kagonya@kagonya.net, anirudha@iitb.ac.in

Abstract. Working within local communities often requires an intimate knowledge of the community in order to enable effective research and design. However, more often than not, it is not easy to substantially engage the local community while conducting research and design, due to constraints and barriers both within and without the project's domain. Common constraints and barriers include navigating socio-cultural, socio-economic and socio-political issues, poor or unfamiliar infrastructure, time and funding constraints, and organizational requirements from both funding agencies and local partners. Moreover, communities may view outsiders with distrust, making collaboration difficult. Consequently, eventual design solutions may be met with resistance or altogether be rejected due to the negative attitude maintained toward the researcher, the funder or the research area. In avoidance of this unfavourable likelihood, researchers and designers have relied on local partners to facilitate their research and design efforts and/or by adapting their processes to suit the local communities. The question posed here however is, how do we ensure rich collaboration with local communities to mutually benefit everyone?

To facilitate discussion and answer these and related questions, a one-day workshop will be conducted to explore best practices in addressing challenges and opportunities associated with collaborating with/in communities. We will explore the value of community-based, community-driven and participatory design and research from different perspectives, and discuss local communication protocols on how to gain trust and manage social change effectively. Best practices and lessons learnt in local community collaboration will be explored, culminating in recommendations for HCI designers and researchers active in the field of local development.

The target audience of this workshop is researcher and practitioners who is interested in, or are already working within local communities.

Read more about the workshop here: http://bit.ly/10pH8sN

Keywords: community collaboration, HCI4D, human-computer interaction, development, co-design, participatory design.

Challenges from the Future: Bridging the Gaps through HCI Education

Lars Oestreicher[1] and Paula Alexandra Silva[2,3]

[1] Dept. of Information Technology, Uppsala University, Uppsala, Sweden
LarsOe@it.uu.se
[2] University of Hawaii at Manoa, Department of Information and Computer Sciences Department, Honolulu, Hawaii
[3] Universidade Portucalense Infante D. Henrique, Departmento de Inovação Ciência e Tecnologia, Porto, Portugal
paulaale@hawaii.edu, palexa@gmail.com

Abstract. The rapid shifting of paradigms within the Human-Computer Interaction field, moving from user interface design to interaction design and onward to user experience design is an issue that strongly affects HCI education and poses additional challenges onto the teacher's agenda. It is difficult for most teachers to keep up with developments that are likely to require a continued redevelopment of the teaching materials.

The workshop will focus on how to prepare HCI Education and Didactics to swiftly adapt to an ever-changing future. The workshop is aimed at HCI teachers, researchers and practitioners who are interested in furthering HCI into the future, and want to be ready to face the challenges triggered by the rapid advancement of technology. The intention is to develop an understanding of these challenges and start a discussion on how to improve the teaching of HCI in a way that it keeps itself up-to-date and ready to address emerging application areas and paradigms shifts.

The dichotomy between teaching about current technology and having a solid foundation for the teaching will be a central issue in the workshop. Is it possible to find alternative ways of teaching that will be less dependent on the current state of technological development? Is it possible to use the traditional knowledge of HCI also within the emerging paradigms of HCI? Are there other developments that can be foreseen and can the associated problems of those be addressed even before the technology has been developed? or is the topic of HCI always going to be about teaching the old reliable knowledge about users and known tools?

The primary objective of the workshop is to, through a constructive discussion between participants, strive towards an understanding of the challenges described above and identify a strategy for a continued development of the teaching of HCI, both as a self-contained and an applied topic.

More information about the workshop can be found at:
http://www.interact2013.org/Workshops

Keywords: HCI Education, HCI Didactics, Technology, Paradigms.

Human Work Interaction Design (HWID): Past History and Future Challenges

Torkil Clemmensen[1], Jose Abdelnour-Nocera[2],
Arminda Lopes[3], and Bengt Sandblad[4]

[1] Copenhagen Business School, Denmark
[2] University of West London
[3] Instituto Politécnico de Castelo Branco, Portugal
[4] Uppsala University, Sweden
tc.itm@cbs.dk, Jose.Abdelnour-Nocera@uwl.ac.uk,
aglopes@ipcb.pt, Bengt.Sandblad@it.uu.se

Abstract. The IFIP 13.6 Human Work Interaction Design (HWID) working group aims at establishing relationships between extensive empirical work-domain studies and interaction design. Today, generic designs are applied to use-situations with very different purposes, as the same social software or games are used for both work and leisure situations. Thus, design shifts from design of a technology to design of various use-situations encompassing the same technological design. We find that there is a need to conceptualize, in HWID models, the relationship between work analysis and design for these new digital realities. The scope of this workshop is to exemplify how HWID approaches translate work analysis to interaction design (and viceversa), and discuss how such understanding can help practitioners and researchers to develop and design digital use situations and digital content. That may entail that we touch upon how theoretical ideas about socio-materiality and socio-technical environments.

In this one-day workshop we aim to make status on the work done within in the IFIP 13.6 Human Work Interaction Design (HWID) approach, and point to future challenges. We invite participants from industry and academia with an interest on empirical work analysis, HCI, interaction design and usability and user experience in work situations and in the workplace. Topics that participants may explore include: Techniques and methods for mapping the relations between work analysis and interaction design; How work analysis can feed into interaction design evaluation; Design cases and case studies of work analysis and in medical and safety critical ICT, enterprise-level systems, e-government services, or mobile devices. The workshop will consolidate - in theoretical HWID models – experiences from empirical case studies of human work analysis and interaction design, and reflect on how these has benefited in enhancing the user experience of a diversity of HWID systems, and provide a set of effective methods and techniques for this purpose. The outcome will be an enhanced HWID framework for studying new digital use situations and digital content.

The workshop will be conducted in an inviting, open and social atmosphere. We aim to provide time for reflection and discussion around each of the accepted papers and cases. For more information, see the workshop website https://sites.google.com/site/interact2013workshophwid/.

Keywords: Human Work Interaction Design, socio-technical, socio-material.

Participatory Design for Persons with Cognitive or Sensory Impairments

Karin Slegers[1], Pieter Duysburgh[2], Niels Hendriks[3], and Helma van Rijn[4]

[1] CUO | Social Spaces – KU Leuven / Minds, Parkstraat 45, bus 3605, 3000 Leuven, Belgium
[2] iMinds – Digital Society, SMIT – VUB, Pleinlaan 9, 1050 Brussels, Belgium
[3] Social Spaces | CUO – MAD-faculty – LUCA – KU Leuven, C-mine 5, 3600 Genk, Belgium
[4] Waag Society, Piet Heinkade 179, 1019 HC Amsterdam, The Netherlands
`Karin.Slegers@soc.kuleuven.be`, `Pieter.Duysburgh@vub.ac.be`, `Niels.Hendriks@khlim.be`, `Helma@Waag.org`

Abstract. When designing for users with impairments involving cognitive or sensory abilities affecting communication (e.g. Autism Spectrum Disorder, Alzheimer's Disease or hearing impairments), common research and design techniques for involving such users might not be usable or need to be adjusted. This full-day workshop aims to 1) bring together researchers and designers who work on participatory design for persons with cognitive or sensory impairments; 2) exchange ideas and experiences regarding participatory design techniques for such persons; 3) identify general principles for participatory design for such persons.

The workshop has a creative, generative character and has four parts:

Introduction. Participants introduce themselves to create an informal atmosphere to share best practices and common problems.

Technique mapping. In small groups, using a template poster, participants discuss participatory techniques they have used with persons with impairments. Next, the groups discuss which elements of each technique they would keep, change, add or delete for future design. Finally, they formulate six 'take home messages'.

Critique. The groups critically review the other groups' take home messages (cf. design critique) by indicating elements they fully agree with, elements they would adjust, and elements they don't agree with at all.

General discussion. Finally, the critiques on the take home messages are presented. Commonalities and differences between the messages are identified, and general principles and key research questions regarding participatory design for persons with impairments are extracted.

The workshop has two follow-up activities: 1) Participants are invited to join a wiki that was set up for sharing information, literature, techniques, etc.; 2) A call for papers for a special issue of the journal CoDesign.

More information on the workshop can be found at the workshop website: http://interact2013impairmentsworkshop.wordpress.com/

Keywords: Cognitive impairments, Sensory impairments, Tools & techniques, User involvement, Participatory design.

Peripheral Interaction: Embedding HCI in Everyday Life

Doris Hausen[1], Saskia Bakker[2], Elise van den Hoven[3,2],
Andreas Butz[1], and Berry Eggen[2]

[1] Human-Computer-Interaction Group, University of Munich (LMU), Germany
{doris.hausen,andreas.butz}@ifi.lmu.de
[2] Industrial Design Department, Eindhoven University of Technology, The Netherlands
{s.bakker,j.h.eggen}@tue.nl
[3] Faculty of Design, Architecture & Building, University of Technology, Sydney, Australia
elise.vandenhoven@uts.edu.au

Abstract. The comparison of actions in the physical world with actions on interactive devices reveals a remarkable difference. In daily life we easily perform several tasks in parallel, for example when drinking coffee while reading, drinking may be in the periphery of the attention. Contrarily, we usually have to focus our attention on each digital device we interact with. In recent years, the concept of interacting with computing technology in the background or periphery of the user's attention is gaining traction. We call this direction Peripheral Interaction, and see it as a very promising approach to fluently embedding the increasing number of interactive devices into our everyday lives.

The workshop is intended to encourage hands-on explorations and discussion about the definition of Peripheral Interaction, its design space and suitable evaluation strategies. Albrecht Schmidt will give a keynote, entitled "Creating Seamless transitions between Central and Peripheral User Interfaces". While the term Peripheral Interaction is not (yet) widely adopted, several design disciplines already address different aspects of the core ideas of Peripheral Interaction (e.g. ambient information systems, ubiquitous computing, implicit interaction, eyes-free interaction, calm technology). We want to sharpen the focus for Peripheral Interaction by offering a platform for exchange of knowledge and community building to establish a network around Peripheral Interaction for further collaboration.

This workshop invites researchers and practitioners from different disciplines (e.g. computer science, interaction design, interactive arts, psychology, cognitive science, product design and social science), to share their experiences with human-computer interaction for the everyday routine, and aims to lay the foundations for a structured exploration of the new interaction paradigm of Peripheral Interaction.

More information about the workshop is available at the workshops website www.peripheralinteraction.com and www.interact2013.org/workshops.

Keywords: peripheral interaction, human attention, trained routines, calm technology, ambient information, interaction design.

Rethinking Universal Accessibility: A Broader Approach Considering the Digital Gap

Julio Abascal[1], Simone Diniz Junqueira Barbosa[2], Andrew Dearden[3], Anirudha Joshi[4], Daniel Orwa Ochieng[5], Gerhard Weber[6], and Panayiotis Zaphiris[7]

[1] Laboratory of HCI for Special Needs. University of the Basque Country, Spain
julio.abascal@ehu.es
[2] Informatics Department. Pontifical Catholic University of Rio de Janeiro, Brazil
simone@inf.puc-rio.br
[3] Sheffield Hallam University, United Kingdom
A.M.Dearden@shu.ac.uk
[4] Industrial Design Centre, IIT Bombay, India
anirudha@iitb.ac.in
[5] School of Computing and Informatics. University of Nairobi, Kenya
dorwa@uonbi.ac.ke
[6] HCI Chair. Technische Universität Dresden
gerhard.weber@tu-dresden.de
[7] Department of Multimedia and Graphic Arts. Cyprus University of Technology, Cyprus
panayiotis.zaphiris@cut.ac.cy

Abstract. "Universal Accessibility" aims at including all people with disabilities in the target population for user-centered design. It is usually focused on the needs of people with disabilities, frequently including elderly people. Current advancements in *Design for all* may require complex equipment, advanced network environments, or computer training. Therefore, these products or services may fall out the reach of many people with disabilities because of economical, technological or formative restrictions. This is a frequent situation in countries where poverty is associated with barriers to obtain assistive technologies and services. This Workshop aims at rethinking the concept of Universal Accessibility in order to find ways to include all excluded people with disabilities. The main objective of this workshop, organized on behalf of the *IFIP WG 13.3 HCI and Disability for HCI*, is to discuss the possibility of broadening the Universal Accessibility concept to include geopolitical and socio-economical circumstances. To this end, attendees will try to find agreed answers to questions such as: Can the Universal Accessibility concept be extended in order to include people with disabilities also affected by other barriers such as poverty, illiteracy, lack of access to advanced technology, etc.? Is it possible to create accessibility guidelines that also consider affordable devices, deprecated technology, cultural issues, or illiteracy? More information about this workshop including a call for participation and post-publication plans can be found in: http://www.egokituz.org/WS/RethinkingUA/

Keywords: Universal Accessibility, Digital gap, Info-exclusion.

Second International Workshop on (Re)Creating Lively Cities through Ambient Technologies: Arts, Culture and Gastronomic Experiences*

Artur Lugmayr[1], Jaz Hee-jeong Choi[2], and Kirralie Houghton[2]

[1] EMMi Lab., Tampere Univ. of Technology (TUT), POB. 541, Finland
[2] Urban Inform.Lab., Queensland Univ. of Techn. (QUT), Kelvin Grove QLD 4059, Australia
lartur@acm.org, {h.choi,kirralie.houghton}@qut.edu.au

Abstract. Digital and interactive technologies are becoming increasingly embedded in everyday lives of people around the world. Application of technologies such as real-time, context-aware, and interactive technologies; augmented and immersive realities; social media; and location-based services has been particularly evident in urban environments where technological and sociocultural infrastructures enable easier deployment and adoption as compared to non-urban areas. There has been growing consumer demand for new forms of experiences enabled through technologies. The focus is on ambient media services, applications, and technologies that promote people's engagement in creating and re-creating liveliness in urban environments, particularly through arts, culture, and gastronomic experiences. The workshop takes a multidisciplinary and future oriented approach, and welcomes participants from diverse disciplinary domains for open discussions about technological, sociocultural, and content-related aspects of ambient media services that support people's engagement in (re)creating their urban environments into a livelier place through art, cultural, and gastronomic experiences*:

Quality of Experience (QoE) and analysis of videos related to art, culture, and gastronomy and urban cities; Ubiquitous environments and interfaces in lively city environments; Intelligent appliances and gadgets supporting art, culture and gastronomy; Multimedia learning for activities around smart city environments: Locative media and context sensor technologies; Artistic, cultural, and gastronomic services; Socio-economic studies, business models, advertising, and marketing; Applied ambient media technologies in city environments.

Read more about the workshop here:
www.tut.fi/emmi/WWW/ameamain/relci2013

Keywords: ambient media, urban experience, usability, urban informatics.

* This workshop description has been replicated from: A. Lugmayr, J.H.-J. Choi, and K. Houghton, "1st International Workshop on (Re)Creating Lievely Cities through Ambient Technologies: Arts, Culture, and Gastronomic Experiences," *Multimedia and Expo Workshops (ICMEW), 2012 IEEE International Conference on*, Melbourne, Australia: 2012.

Urban Agriculture: A *Growing Field* of Research

Peter Lyle[1], Jaz Hee-jeong Choi[1], Shaun Lawson[2], Christopher Lueg[3], Alan Chamberlain[4], Marcus Foth[1], Anna Meroni[5], and Margot Brereton[1]

[1] Queensland University of Technology, Brisbane, Australia
[2] Lincoln University, Lincoln, UK
[3] University of Tasmania, Hobart, Australia
[4] Nottingham University, Nottingham, UK
[5] Politecnico di Milano, Milan, Italy
{p.lyle,h.choi}@qut.edu.au, slawson@lincoln.ac.uk, Christopher.Lueg@utas.edu.au, Alan.Chamberlain@Nottingham.ac.uk, m.foth@qut.edu.au, anna.meroni@polimi.it, m.brereton@qut.edu.au

Abstract. Growing food presents diverse challenges and opportunities within the urban environment. As cities develop, population density rises, land prices rise, and the opportunity to use land for traditional farming and gardening diminishes. Counter to this trend there are a growing number of community gardens, city farms, guerrilla gardening, rooftop and vertical gardens, pot plants, windowsill herbs, and other balcony or backyard gardens cropping up in cities, all with a purpose to produce food. This workshop brings together practitioners and researchers in the field of urban agriculture and Human-Computer Interaction to explore and opportunities for technology design to support different forms of growing practice and foster local food production in cities.

This 1-day workshop will serve as an active forum for researchers and practitioners across various fields including, but not limited to, agriculture and gardening, education, urban planning, human-computer interaction, and community engagement. This workshop has three distinct points of focus: i) Individual and small-scale gardening and food production, and how to connect like-minded people who are involved in these practices to share their knowledge ii) Communities involved in urban agriculture, either through community gardens, city farms, or grassroots movements, often dependent on volunteer participation, providing the challenge of managing limited resources iii) Environmental and sociocultural sustainability through urban agriculture..

The participants will have an opportunity to present their own work. This will be followed by a visit to a nearby city farm, which will provide a local context for a group design exercise. Finally the workshop will conclude with panel discussions to review opportunities for further research and collaborations beyond the conference.

For more information, please visit the workshop website, at http://www.urbaninformatics.net/resources/interact2013cfp/

Keywords: Urban agriculture, Communities, City farms, Food security, Environmental sustainability.

Walking for Data

Mads Bødker[1], David Browning[2], and Nina Dam Meinhardt[1]

[1] Department of IT Management, Copenhagen Business School
{mb.itm,ndm.itm}@cbs.dk
[2] School of Natural Resources and Tourism, Polytechnic of Namibia
dbrowning@polytechnic.edu.na

Abstract. HCI practitioners are often involved in developing mobile applications that pay little heed to the experiences of mobility and place. Although applications commonly utilise location data to deliver contextual information, (e.g. map apps, tourist guides and sports record keeping apps), little consideration is given to how the experience of being mobile and the attendant place making could be used in design. Indeed, bodily movement whilst using a mobile device is often considered problematic! Researchers' sensitivity to the interactions mobility affords is often restricted by difficulties in collecting data whilst moving about, for example, whilst walking, Yet, as Bidwell illustrates, the importance for design of going beyond devices and methods that "constrain our world to points" is overdue. Researchers are thus driven to develop practical methods for collecting such data.

A one-day workshop, intended for HCI practitioners and interaction designers, will be conducted to facilitate discussion of the theoretical issues, and to explore mobile data collection and analysis of the resultant place making through what we have labelled 'walking methods'. Video, the obvious choice for recording mobility, tends to compound the problem being both difficult to handle because of the amount of information it contains and by providing too narrow a bandwidth to render a satisfactory simile of the experience. Indeed, watching a video recording of mobile interaction is often a pale, non-immersive reflection of the experience. Yet such material affords cued-recall debriefing. By itself this is not sufficient, but it is useful in the process of interacting with the recorded data to develop artifacts that can inform and shape a design space. So, a designer's role is not necessarily to walk the walk but rather to draw inspiration and to tease out innovation from such artifacts. Based on the day's activities, the workshop will culminate in a discussion of how mobile interactions might yield material applicable to design.

Please visit **http://walkingfordata.tumblr.com** for additional information.

Keywords: walking, place, ethnography, design, methods.

Author Index