Bilge Mutlu Christoph Bartneck
Jaap Ham Vanessa Evers
Takayuki Kanda (Eds.)

LNAI 7072

Social Robotics

Third International Conference, ICSR 2011
Amsterdam, The Netherlands, November 2011
Proceedings

Springer

Lecture Notes in Artificial Intelligence 7072

Subseries of Lecture Notes in Computer Science

Bilge Mutlu Christoph Bartneck
Jaap Ham Vanessa Evers Takayuki Kanda (Eds.)

Social Robotics

Third International Conference, ICSR 2011
Amsterdam, The Netherlands, November 24-25, 2011
Proceedings

Springer

Series Editors

Randy Goebel, University of Alberta, Edmonton, Canada
Jörg Siekmann, University of Saarland, Saarbrücken, Germany
Wolfgang Wahlster, DFKI and University of Saarland, Saarbrücken, Germany

Volume Editors

Bilge Mutlu
University of Wisconsin-Madison
Madison, WI 53706-1685, USA
E-mail: bilge@cs.wisc.edu

Christoph Bartneck
University of Canterbury
Christchurch 8140, New Zealand
E-mail: christoph.bartneck@canterbury.ac.nz

Jaap Ham
Eindhoven University of Technology
5600 MB Eindhoven, The Netherlands
E-mail: j.r.c.ham@tue.nl

Vanessa Evers
University of Twente
7500 AE Enschede, The Netherlands
E-mail: v.evers@utwente.nl

Takayuki Kanda
Advanced Telecommunications
Research Institute International
Keihanna Science City
Kyoto 619-0288, Japan
E-mail: kanda@atr.jp

ISSN 0302-9743 e-ISSN 1611-3349
ISBN 978-3-642-25503-8 e-ISBN 978-3-642-25504-5
DOI 10.1007/978-3-642-25504-5
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2011941491

CR Subject Classification (1998): I.2, C.2.4, I.2.11, H.5.2, J.4, I.2.10

LNCS Sublibrary: SL 7 – Artificial Intelligence

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)

Message from the General Chairs

In this volume, you will find the papers presented at the Third International Conference on Social Robotics (ICSR), held during November 24–25, 2011, in Amsterdam, The Netherlands. In the new and rapidly growing and evolving research area of social robotics, building a community that is collegial, supportive, and constructive is crucial for fostering the scientific qualities needed to answer the questions that this strongly interdisciplinary field poses. The diversity of backgrounds and the sheer number of the Chairs involved in organizing this conference characterizes that enterprise. Likewise, the diversity of the papers in these proceedings and of the research discussed at the conference is an indication of the growing interest in social robotics research from a multitude of perspectives.

ICSR 2011 built strongly on the earlier ICSR conferences (Korea, 2009; Singapore, 2010), and was very much in debt to the huge efforts of the Standing Committee headed by Shuzhi Sam Ge. We thank all Organizing Chairs of ICSR 2011 for their tremendous efforts in making this conference a success; not only in its scientific output, but also in connecting researchers from all over the globe.

November 2011

Jaap Ham
Vanessa Evers
Takayuki Kanda

Preface

The Program Chairs proudly present the proceedings of the Third International Conference on Social Robotics. We were extremely pleased to see the growing interest in the conference and the extraordinary effort by our co-organizers and the research community to create a high-quality venue for publishing and sharing scientific research in social robotics.

The main focus of our work as Program Chairs has been to nurture high-quality scientific contributions to the conference by establishing a fair and rigorous review process. We sought to strictly adhere to submission deadlines and introduce a two-stage review process and handling of papers by meta-reviewers—experts and leaders in social robotics made up of faculty members and senior researchers. Fifty-one papers were initially submitted to the conference. Each paper was managed by one of 20 meta-reviewers who assigned a minimum of two external reviewers to the paper. A total of 85 reviewers from the social robotics community contributed to the review process as external reviewers. After the reviewers completed their evaluations, the meta-reviewers summarized them in light of their expert opinion and made their recommendation. Twelve papers were accepted and 13 were conditionally accepted.

The conditional papers entered the second stage of the review process, which we refer to as the "shepherded process." The meta-reviewers took a more active role, working with the authors to improve the papers and ensuring that the conditions for acceptance were fully satisfied. Eleven out of 13 papers were approved through this process. This review process resulted in a total of 23 outstanding papers that are included in the technical program of the conference, resulting in a 45% acceptance rate in the full-paper track.

We acknowledge that some early research work might not yet have been mature enough to be included in the conference technical program. The Works-In-Progress (WIP) Track offers authors the opportunity to discuss early results of their work with the community. Not archiving their papers allows the authors to keep the copyright of the papers. We hope that the feedback they received during the conference contributes to the completion of their work and the authors submit their completed work to next year's conference.

We would like to thank all the Chairs, Co-chairs, meta-reviewers, and reviewers who made such an extraordinary effort to bring this conference to life.

November 2011

Christoph Bartneck
Bilge Mutlu

Organization

The International Conference on Social Robotics brings researchers and practitioners together to report and discuss the latest progress in the field of social robotics. The conference focuses on the interaction between humans and robots and the integration of robots into our society. The inaugural conference was held in Incheon, Korea (2009), and after the very successful ICSR 2010 in Singapore, we were proud to invite participants to the intriguing city of Amsterdam, The Netherlands.

The theme of the 2011 conference was "Alive!" It expresses the vitality of the social robotics research, paying particular attention to the development of robots that appear increasingly social – the point that people perceive them to be alive. The conference aims to foster discussion on the development of computational models, robotic embodiments, and behavior that enable robots to act socially and the impact that social robots have on people and their social and physical environment.

Standing Committee

Shuzhi Sam Ge	National University of Singapore/University of Electronic Science and Technology of China
Maja Mataric	University of South California, USA
Haizhou Li	A*Star, Singapore
Jong-Hwan Kim	KAIST, Korea

General Chairs

Vanessa Evers	University of Amsterdam, The Netherlands
Jaap Ham	Eindhoven University of Technology, The Netherlands
Takayuki Kanda	ATR IRC, Japan

Program Chairs

Christoph Bartneck	University of Canterbury, New Zealand
Bilge Mutlu	University of Wisconsin-Madison, USA

Program Committee

Selma Sabanovic	Indiana University, USA
Takanori Komatsu	Shinshu University, Japan
Bruce MacDonald	Auckland University, New Zealand

Kolja Kuehnlenz	Technical University of Munich, Germany
Michio Okada	Toyohashi University of Technology, Japan
Kai O. Arras	University of Freiburg, Germany
Sarah Perez-Kriz	University of Washington, USA
Aaron Steinfeld	CMU, USA
Manfred Tscheligi	University of Salzburg, Austria
Yuichiro Yoshikawa	Osaka University, Japan
Myung-Suk Kim	KAIST, Korea
Leila Takayama	Willow Garage, Inc., USA
Cindy Bethel	Yale University, USA
Dylan Glass	ATR, Japan
Christian Becker-Asano	University of Freiburg, Germany
Masayuki Kanbara	NAIST, Japan
Tony Belpaeme	University of Plymouth, UK
XiaoQi Chen	University of Canterbury, New Zealand
Adriana Tapus	ENSTA, France
Miguel A. Salichs	Carlos III University, Spain
Yeow Kee Tan	A*Star, Singapore

Tutorial Chairs

Martin Saerbeck	A*STAR, Singapore
See Swee Lan	A*STAR, Singapore

Robot Design Competition Chairs

Adrian Tay Hwang Jian	A*STAR, Singapore
Henrik Scharfe	Aalborg University, Denmark
Raymond Cuijpers	Eindhoven University of Technology, The Netherlands

Publication Chairs

Dana Kulić	University of Waterloo, Canada
John-John Cabibihan	National University of Singapore, Singapore

Work-In-Progress Chairs

Ravindra De Silva	Toyohashi University of Technology, Japan
Dennis Reidsma	Twente University of Technology, The Netherlands

Exhibition Chair

Astrid Weiss	University of Salzburg, Austria

Finance Chair

Birna van Riemsdijk — TU Delft, The Netherlands

Website Chair

David van der Pol — Eindhoven University of Technology, The Netherlands

Reviewers

Eun Ho Kim
Ross Mead
Astrid Weiss
Tatsuya Nomura
Bill Smart
Brian Scassellati
Sonya Kwak
Hirotaka Osawa
Paul Bremner
Christopher Ford
Frederic Delaunay
Antoine Hiolle
Jeong-gun Choi
Yunkyung Kim
Kwangsu Cho
Jeonghye Han
Thierry Chaminade
Anna-Lisa Vollmer
Hatice Kose-Bagci
Peter Asaro
Henny Admoni
Christian Smith
Michael Mistry
Tony Kuo
Chandan Datta
Geoffrey Biggs
Maria Malfaz
Christian Schlegel
Boris Duran
Fumihide Tanaka
Yoshio Matsumoto
Kohei Ogawa
Tobias Rehrl
Nina Riether
Candace Sidner
Vasant Srinivasan
Xiaojuan Ma
Mark Coulton
Manuel Birlo
Weihua Sheng
Daniel Grollman
Yugo Takeuchi
Lara Lammer
Cyrill Stachnis
Gabe Sibley
Stephan Weiss
Raj Madhavan
Dirk Holz
Barbara Gonsior
Claudia Muhl
Nate Koenig
Philipp Baer
Nicole Mirnig
Matthew Lee
Antonia Glaser
Yohei Kamo
Solace Shen
John Porter
Masahide Yuasa
Kazuhiko Shinozawa
Rick Dale
Ryan Calo
Kazuki Kobayashi
Hisashi Ishihara
Louis-Philippe Demers
Yoichi Matsuyama
Christoph Bartneck
Stefan Sosnowski
Kotaro Funakoshi
Elizabeth Broadbent
Nejat Goldie
Masahiro Shiomi
Manja Lohse
David Feil-Seifer
Frank Hegel
Sonia Chernova
Joachim de Greeff
Kirsten Bergmann
Ravindra De Silva
Astrid von der Pütten
Paul Baxter
Raquel Ros Espinoza
Michihiro Shimada
Stefan Scherer

Sponsoring Institutions

Maxon Motors

Cyberbotics

Metra Labs

Aldebaran Robotics

Table of Contents

Social Interaction with Robots

Nonverbal Interaction with Social Robots

Robots in Society

Social Robots in Education

Affective Interaction with Social Robots

Robots in the Home

Interaction Scenarios for HRI in Public Space

Jakub Złotowski, Astrid Weiss, and Manfred Tscheligi

ICT&S Center,
University of Salzburg,
Sigmund-Haffner-Gasse 18,
5020 Salzburg, Austria
{jakub.zlotowski,astrid.weiss,manfred.tscheligi}@sbg.ac.at

Abstract. Social acceptance of robots in public space implicitly requires a user-centered approach in the development phase of a robotic system. One of the most popular tools used by interaction designers in the field of Human-Computer Interaction (HCI) are the persona-based scenarios. Recently, they find their way more and more into HRI research. This paper describes how specifically designed human-human studies can support the modeling process of interaction scenarios in HRI. As part of the Interactive Urban Robot (IURO) project, we analyzed three human-human studies to better understand human selection process of interaction partners and an approached person's subsequent actions in itinerary request situations. This empirical data was used to create personas and scenarios for HRI in public space.

Keywords: scenarios, interaction design, user-centered design, social acceptance, public space.

1 Introduction

The rapid development of robotic systems, which we can observe in recent years, allowed researchers to investigate HRI in places other than the prevailing factory settings. Robots have been employed in shopping malls [4], train stations [3], schools [13], streets [14] and museums [10]. In addition to entering new human environments, the design of HRI recently started shifting more and more from being solely technologically-driven towards a user-centered approach. In accordance with HCI design guidelines, user needs and interaction scenarios are taken into account in the early stages of a social robot's design process. This approach helps to take into account specific user requirements in the development of a robot's appearance and behavior. The first results of long-term HRI studies based on a user-centered design approach [6] and ethnographic approaches "in the wild" [12] are now available, indicating the relevance of in-depth understanding of human behaviors and needs for successful HRI. Similarly, the first methodological approaches are developed to enable evaluation of HRI [7]. Moreover, Steinfeld et al. [11] presented a promising approach of simulating the human, as human modeling permits predictions of adaptive HRI.

B. Mutlu et al. (Eds.): ICSR 2011, LNAI 7072, pp. 1–10, 2011.

In line with this recent trend, we present an approach in which users are involved from the very first stages of a robot's development process. This paper describes how to create scenarios, based on human-human studies, for HRI in situations where current robot's development makes it difficult to efficiently gather data about the system's potential users. Furthermore, to aid the creation process of HRI scenarios, personas are used as a support tool to make the scenarios vivid and lifelike. The work presented in this paper is part of EU FP7 IURO project[1], which develops a robot that is capable of navigating in densely populated human environments using only information obtained from encountered pedestrians. This project is a follow-up project to the nationally funded pilot study of the Autonomous City Explorer (ACE) project. The ACE robot interacted via gesture and touch screen input, and speech and image output, whereas the interaction was controlled by a finite state machine. The interaction started with the robot greeting a pedestrian and the itinerary request. The robot then asked the pedestrian to point in the direction of the designated goal for establishing a reference point for the itinerary request. Then pedestrians could indicate further directions via buttons on the touch screen. During the interaction ACE builds a topological route graph from the presented route information. At the end, the robot thanked the pedestrian and followed the route graph (see [14] for further details). The goal for the IURO robot is to use speech and gesture as input and output.

Our goal for the scenario approach was to use this user-centered design technique in order to communicate the idea of how the interaction with IURO should look like in the future to project partners from various disciplines. The scenarios should help all project partners to create a consistent mental model of the interaction with IURO. By the means of vivid integrated personas, scenarios underline the relevance of an in-depth understanding of humans' behaviors and the needs of user-centered design for successful short-term HRI in public space. Thereby, the narrative stories support the continuous consideration of potential user behaviors during the development of the robotic platform. Moreover, the scenarios presented in this paper should also enable other HRI researchers who work on similar scenarios to take the user perspective into account.

The paper is structured as follows, we firstly present current practice from HCI and HRI regarding a persona-based scenario creation process. Secondly, the idea behind using human-human study-based scenarios is introduced, followed by a presentation of the developed personas and scenarios. At that point, the benefits of these scenarios, which we believe can be also applicable in other projects, are discussed. Finally, conclusions and limitations of the proposed approach are presented.

2 Scenarios

Scenarios are narrative stories consisting of one or more actors with goals and various objects they use in order to achieve these goals. In HCI scenarios are

[1] `http://www.iuro-project.eu/`

used at various stages of the product development cycle, such as for the problem definition, analysis, design and evaluation of systems [9]. They help to communicate ideas between team members as well as between designers and clients. The narrative aspect supports the involved parties in creating common mental models and build a shared system vision. Furthermore, they help to focus on users and their needs instead of concentrating on the product. Finally, scenarios help to avoid the problem of fixating on the first solution as, while being enough concrete to support common understanding, they are also flexible to support numerous solutions.

Usually the actors used in scenarios are called personas. According to Cooper [1] personas are "based upon behavior patterns we observe during the course of the *research* phase, which we then formalize in the *modeling* phase". The main goal of personas is to ensure that the product being developed is designed for concrete users rather than an abstract, non existing "average user". Often, more than one persona is created in order to address the whole spectrum of the target group.

Persona-based scenarios are popular design tools in HCI. In recent years they paved their way to HRI. Robins et al. [9] presented a general methodological approach for scenario development in HRI research in the context of robot assisted children play. They showed how a panel of experts can support the validation of literature review-based scenarios. On the other hand, there are areas where the novelty of robotic systems complicates the data collection of the target user group. In the IURO project we explore a general scenario where the robot has no access to a map and needs to navigate from one place in a city to another using only information acquired from pedestrians. This is an alternative scenario for situations when a robot is unable to accomplish a given task using only predefined information.

Therefore, when navigating through urban environments, IURO is faced with different kinds of situations that require different kinds of navigation behaviors. It must be able to explore the environment, follow a sidewalk, safely cross a street, approach a person or follow a certain direction. Consequently, the robot must be able to find a human and initialize the interaction, whereby successful speech interaction and gesture recognition pose sever challenges, due to background or traffic noise at heavily frequented public places and changing light conditions. Since currently there are no robots present in public spaces, it is a completely new situation for people and a potential target group cannot be easily identified. Thus, we created HRI scenarios based on human-human studies as cost effective method to be used in the early phases of the robot's development.

3 Using Human-Human Studies in User-Centered HRI

In comparison with other computer technologies, humanoid robots encourage more anthropomorphic mental models [5]. As a result, we decided to conduct a series of human-human studies in order to obtain insights on potential HRI in public space. In total, three studies were conducted in the streets of Salzburg

(Austria): Study 1 - Itinerary Requests (n=40), Study 2 - Pedestrian Selection (n=47), Study 3 - Interaction Reasons (n=10). Additionally, footage from a field trial with the Autonomous City Explorer (ACE) robot was analyzed to bridge the gap between differences in HHI and HRI (n=52). The studies were based on "ecological", semi-experimental encounters, initiated either by researchers or recruited participants (all of them living in Salzburg) who asked people for directions to a specified pharmacy[2]. Questionnaires, interviews and the retrospective think-aloud method (based on recorded interaction videos) were used to find the influencing communication factors for successful conversation and selection criteria for choosing an interaction partner. For more details on these studies please refer to [15].

The data corpus, which in total consisted of 149 itinerary request situations, was then used to create personas for HRI scenarios. Furthermore, the results of these studies had direct impact on the scenario creation process, as they helped to define positive and negative scenarios, based on the analysis of the successfulness (receiving correct directions) of human actions. The following sections explain in detail how the personas and scenarios were created and present examples developed for the needs of the IURO project.

4 Deriving Personas from Human-Human Data

Following the guidelines of Dix et al. [2] for the scenario creation process, we first developed personas. While it is a common practice to use qualitative data for the development of personas, we decided to also include quantitative data as proposed by McGinn and Kotamraju [8]. This method helped us to identify the potential user group of the IURO robot as using solely qualitative data would have been ineffective and potentially misleading due to the futuristic scenario explored in this project. Since people are experts in asking pedestrians for the way, we assumed that IURO should use similar behavior strategies. The source material for the persona development consisted of videos, dialogue transcripts, and questionnaire data, whereas the qualitative data was annotated and interpreted by means of content analysis with the NVivo software tool.

4.1 The Method

The process of persona creation started with the identification of key demographic aspects of the selected pedestrians. We analyzed their age range, profession, education and language skills. This information was then enriched by the data obtained during interviews where we asked participants why they approached specific pedestrians. Not surprisingly, we found that one of the most important factors, which impacts the successfulness of the interaction, was whether the encountered person was a local or not.

[2] Only in the Study 2 - Pedestrian Selection participants knew what the experiment was about, in the other two studies they were informed during the debriefing phase.

In addition, the data also showed that people base their judgment about a person being an inhabitant on her movement (such as goal-directed walking), physical appearance, possession of local items or behavior (people who are working in public space, such as a waiter or salesman). At that point we were able to draw vivid personas including their demographic data, physical appearance and even daily habits. Based on the successfulness of the interaction we identified a primary and a secondary persona. However, almost straight away it became clear that people, despite being relatively accustomed to asking others for help in public space, sometimes selected pedestrians who were unable to help them (in most cases due to the fact that the city center of Salzburg is heavily frequented by tourists). It led us to the decision to include a negative persona as IURO should be able to recover from encountering such a situation.

4.2 The Personas

The primary persona is Mary Bishop, a saleswoman at a tourist shop, who lived all her life in the city. The secondary persona is Peter Cooper, a website designer, who moved to the city. The third persona is the negative one, a 20 years old student from Spain who visits the city for the first time. All three personas were situated in the context of the German city Munich, which is highly comparable to Salzburg (Austria), as Munich will be the site for all forthcoming interaction studies with IURO. Two of them are exemplary presented in the following.

Primary persona
Mary Bishop

- 48 years
- Tourist shop saleswoman
- Lived in Munich all her life
- Uses Internet mainly to check emails and find information about local concerts
- No previous experience with robots
- Native German speaker

Goals: Feel knowledgeable, run shop effectively, socialize with people

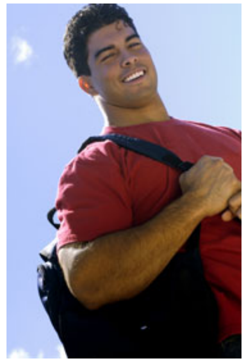
Credits: http://www.gettyimages.at/detail/EB3691-001/Riser

A day from her life: Mary always wakes up at 7am. After taking a shower and eating breakfast she leaves at 8am to a family-owned tourist shop, located in the Old Town of Munich. She goes there by bus and then walks 500 meters from the bus stop to the shop. She arrives around 20 minutes before the shop opens, to make sure that everything is ready before 9am - the official opening time. Despite having more than enough time, Mary is always worried that she will not prepare everything on time as she is the only salesperson working in the shop. Usually, in the morning there are not that many tourists visiting her shop so when there is nobody inside and the weather is good she likes to go outside her shop and stand next to the entrance. She has a close friend working in a restaurant located just next door and often, when both do not have customers, they talk about everyday life matters and plans for forthcoming Jazz concerts.

Negative persona
Juan Rodrigues

- 20 years
- Student and tourist
- Comes from Spain and is visiting Munich for the first time
- Uses computers daily as a tool for studies, and the Internet to plan trips and for social networking
- No previous experience with robots. Has a touch screen mobile phone
- He is native Spanish speaker and speaks English fluently. However, he does not speak German

Goals: Have fun, learn about other cultures and see new places



Credits:
http://firstnyfcu.files.wordpress.com/2008/01/
college-student.jpg

A day from his life: Juan studies at the University of Madrid (Spain). He is a born adventurer, whose life is traveling. He spends every possible break during the academic year on tours around Europe. He is even ready to skip some of the classes, if there is some special event in a city he would like to visit. His friends are not that passionate about traveling so, while during summer vacations he travels with a bunch of friends, during the academic year he is sightseeing usually alone. Juan has never been to Munich. However, it is October and he always wanted to experience Oktoberfest, so he decided to get a week off from school and came to Munich.

These newly created personas facilitated discussions between the project members as they helped to emphasize which characteristics of a person increase the probability of providing the requested information. Furthermore, it became clear what physical aspects should be analyzed by IURO while choosing an interaction partner, which was important information for the team working on the robot's vision. In the second phase of user-centered design taken in the project, we have incorporated the personas in scenarios.

5 Deriving Scenarios from Human-Human Data

Already knowing which people should be approached, we focused now on how to approach them; how the interaction is initiated and how do they ask for directions. Similarly to personas, our scenarios were based on the previously described human-human studies.

5.1 The Method

In the first phase, we identified the side people tend to approach pedestrians from and what nonverbal cues they send in order to signal their intentions. Moreover, we were interested in learning more about modalities and oral information which people provide while describing landmarks and directions. The analysis also involved data on how the reported shortage of time of pedestrians impacts their willingness for helping. At that point, we then integrated personas to create vivid and lifelike scenarios which depicted crucial factors for successful and unsuccessful interaction of a social robot in public space. The behavior of IURO in these scenarios mimics human behavior as it was observed in public space. In total, two successful scenarios (interaction with a saleswoman and interaction with multiple users) and one unsuccessful scenario (interaction with a foreigner) were developed. Two of them are exemplary presented in the following subsection.

5.2 The Scenarios

Scenario 1 - Saleswoman It is a sunny day in Munich. There are many tourists in the city, but as it is lunch time most of them are currently eating in restaurants. As a result there is nobody in Mary's tourist shop at the moment. She never loses this kind of opportunity to stand outside for some minutes. As she is looking at the square in front of her shop, she notices the IURO robot moving towards her. It approaches her and asks for directions to the nearest pharmacy. Mary is slightly surprised at first, but as she knows even smaller streets of the Old Town and is used to give directions to tourists, she tells IURO to go 50 meters straight on towards Maximilianplatz and turn right in Maxburgerstrasse. As she is saying this, she also points with her hand towards the direction. After hearing feedback from the robot that it understood the given instructions she adds that a pharmacy will be on the right hand side of the street. IURO repeats the instructions as it understood them and as they are correct Mary says: "That's right". The robot thanks her and starts navigating towards the pointed direction. Mary is thinking about this surprising encounter, but at the same time feels happy as always when she was able to help someone.

Scenario 2 - Meeting a foreigner. Juan woke up very late today after celebrating the beginning of Oktoberfest last evening. Since he does not have much time to spend in Munich, he decides to go to the city straight away and eat brunch in one of the many cafes. As soon as he arrives to the Old Town he is surrounded by a crowd of other people. He is not sure where to go, so he decides to wander around for a few minutes until he sees a suitable place for eating. Suddenly, he bumps into the IURO robot, which is moving frontally towards him. As the robot arrives closer, it looks at Juan and says something. Juan stops. The robot also stops around 1 meter in front of Juan. He asks the robot to repeat what it said, but it speaks German, which Juan does not understand. He is interested to know what the robot is saying and notices a touch screen display, but the text is also displayed in German only. After little success with the interaction he hears familiar word in German: "Danke" (Thank you), which he interprets as the robot's intention for ending the conversation. Both, the robot and Juan start moving in opposite directions and the former approaches the next passing by person.

6 Benefits of Using Human-Human Study-Based Scenarios in HRI

The personas and scenarios presented in this paper are currently used in the IURO project. They have already brought several benefits to the project members. In the first place, human-human studies were much faster and cheaper to conduct compared to using the robot in an experimental setup. Furthermore, it was possible to gain first insights on the context of human interaction in public space when the development of the robot was still in the early phase and it was not feasible at that time to perform any experiments with it.

Moreover, as the IURO project involves several partners, the scenarios helped them to create a consistent mental model of how the interaction with IURO should look like and which aspects are crucial during the technical development. This is especially important for projects, which include different organizations focused on technical and evaluational aspects from both an industry and academia point of view. In our project, the created personas helped the robot's vision developers in working on a system which will be able to analyze human physical appearance while selecting an interaction partner. In addition, the scenarios emphasized in a narrative way the need for a social robot in public space to perceive various modalities, as people tended to use both verbal and non-verbal communication while explaining directions. More details on design recommendations and guidelines for the communication and approach behavior derived from the three HHI studies can be found in [15]. Furthermore, the scenarios support the development of the speech system for IURO by identifying which words should be included in the robot's dictionary for describing landmarks and distances. It is especially important in the context of public space, where there is a lot of noise and speech recognition will be hampered. Finally, these scenarios were used in the development of path planning for the robot.

We believe that the proposed methodology for creating HRI scenarios can benefit the HRI community. Furthermore, the presented personas and scenarios can provide for a better understanding of the context of social robots in public space. Deriving scenarios from human-human studies has proved to be a cost-effective method by providing viable results for the development of a robotic system. It is especially well-suited in the early phases of the user-centered design cycle.

7 Conclusions

This paper described how using data from specifically designed human-human studies can aid the creation process of HRI scenarios in public space. This approach did not only help creating personas, but also provided insights on the context of the interaction and supported IURO's behavior modeling. This user-centered approach provided and still provides benefits in the design and ongoing development of the robot platform. Moreover, it ensures that the robot's behavior will be as natural as possible for humans.

As the long-term goal of the IURO project is to enable sustainable and complex short-term interaction in public space with a robot, we really need to consider scenarios that go beyond the willingness to help a robot due to its novelty effect. As the videos of the tweenbots project[3] show, the novelty effect of affective artifacts in public space provoke a helping behavior, but how should the behavior look like in a future in which 50 service robots are moving around in the city center and ask pedestrians for information?

We do not want to suggest conducting human-human studies in order to obtain data used in HRI design process at all times. In fact, when it is possible to collect the data about actual users of a robotic system that should be the preferred method (thus we linked our data in the scenario creation process with the results of the previous ACE study). Furthermore, it is necessary to validate interaction scenarios based on human-human studies in the later phases of the system development, when it is feasible to utilize it in a natural human environment. Nevertheless, when user data is difficult to collect and interaction scenarios for a robotic system are hard to define, conducting human-human studies can be a successful and cost-effective method for first insight into an interaction context, and it can aid a user-centered interaction design process in the early phases.

References

1. Cooper, A.: About Face 3.0: The Essentials of Interaction Design. John Wiley & Sons, Inc., New York (2007)
2. Dix, A., Finlay, J., Abowd, G.: Human-computer interaction. Prentice Hall (2004)
3. Hayashi, K., Sakamoto, D., Kanda, T., Shiomi, M., Koizumi, S., Ishiguro, H., Ogasawara, T., Hagita, N.: Humanoid robots as a passive-social medium: a field experiment at a train station. In: Proceedings of the ACM/IEEE International Conference on Human-robot Interaction, HRI 2007, pp. 137–144. ACM, New York (2007)
4. Kanda, T., Glas, D.F., Shiomi, M., Ishiguro, H., Hagita, N.: Who will be the customer?: a social robot that anticipates people's behavior from their trajectories. In: Proceedings of the 10th International Conference on Ubiquitous Computing, UbiComp 2008, pp. 380–389. ACM, New York (2008)
5. Kiesler, S., Hinds, P.: Introduction to this special issue on human-robot interaction. Hum.-Comput. Interact. 19, 1–8 (2004)
6. Lee, M.K., Forlizzi, J., Rybski, P.E., Crabbe, F., Chung, W., Finkle, J., Glaser, E., Kiesler, S.: The snackbot: documenting the design of a robot for long-term human-robot interaction. In: Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction, HRI 2009, pp. 7–14. ACM, New York (2009)
7. Lohse, M., Hanheide, M., Rohlfing, K.J., Sagerer, G.: Systemic interaction analysis (sina) in hri. In: Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction, HRI 2009, pp. 93–100. ACM, New York (2009)

[3] `http://www.tweenbots.com/`

8. McGinn, J.J., Kotamraju, N.: Data-driven persona development. In: Proceeding of the Twenty-sixth Annual SIGCHI Conference on Human Factors in Computing Systems, CHI 2008, pp. 1521–1524. ACM, New York (2008)
9. Robins, B., Ferrari, E., Dautenhahn, K.: Developing scenarios for robot assisted play. In: The 17th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN 2008, pp. 180–186 (2008)
10. Shiomi, M., Kanda, T., Ishiguro, H., Hagita, N.: Communication Robots in Real Environments. In: Humanoid Robots, Human-like Machines, pp. 567–576. InTech (2007)
11. Steinfeld, A., Jenkins, O.C., Scassellati, B.: The oz of wizard: simulating the human for interaction research. In: Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction, HRI 2009, pp. 101–108. ACM, New York (2009)
12. Sung, J., Christensen, H.I., Grinter, R.E.: Robots in the wild: understanding long-term use. In: Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction, HRI 2009, pp. 45–52. ACM, New York (2009)
13. Tanaka, F., Cicourel, A., Movellan, J.: Socialization between toddlers and robots at an early childhood education center. Proceedings of the National Academy of Sciences 104(46), 17954 (2007)
14. Weiss, A., Igelsböck, J., Tscheligi, M., Bauer, A., Kühnlenz, K., Wollherr, D., Buss, M.: Robots asking for directions: the willingness of passers-by to support robots. In: Proceeding of the 5th ACM/IEEE International Conference on Human-robot Interaction, HRI 2010, pp. 23–30. ACM, New York (2010)
15. Weiss, A., Mirnig, N., Buchner, R., Förster, F., Tscheligi, M.: Transferring Human-human Interaction Studies to HRI Scenarios in Public Space. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) INTERACT 2011, Part II. LNCS, vol. 6947, pp. 230–247. Springer, Heidelberg (2011)

MAWARI: A Social Interface to Reduce the Workload of the Conversation

Yuta Yoshiike, P. Ravindra S. De Silva, and Michio Okada

Interactions and Communication Design Lab,
Toyohashi University of Technology,
Toyohashi 441-8580, Japan
yoshiike@icd.cs.tut.ac.jp,
{ravi,okada}@tut.jp
http://www.icd.cs.tut.ac.jp/en/profile.html

Abstract. In this paper, we propose a MAWARI-based social interface as an interactive social medium to broadcast information (e.g. news, etc.). The interface consists of three sociable creatures (MAWARIs) and is designed with minimalism designing concepts. MAWARI is a small scale robot that has only body gestures to express (or interact) its attractive social cues. This helps to reduce the observing workload of the user, because during the conversation the creature just uses vocal interactions mixed with attractive body gestures. The proposed interface is capable of behaving in two kinds of states: behavior as a passive social interface and also as an interactive social interface in order to reduce the conversation workload of the participant.

Keywords: Minimalism designing, passive social interface, multiparty-discourse, conversational workload.

1 Introduction

Human-computer interaction research has emerged to expend robots as an information service medium [14][10]. In these applications, the context is preserved with one-to-one communication (between a human and robot) persisting until the user is gratified with the information obtained from the robot. In the scenario of one-to-one communication, both parties (human and robot) need to exert much effort to do answers/questions to satisfy one another to understand their expectations. This kind of conversation induced both parties in having a workload for the conversation, because to continue the conversation both parties must be compelled to respond chronologically to the communication. As we know, body gestures, facial expression, eye gaze, and turn-taking management are also essential factors to continue efficacious conversations among those involved [3]. Also, affective social cues, personal perceptions, and group formation are partly involved in establishing effective communication, but there is still unknown and insufficient evidence to understand the scale of the involvement [15]. Since questions still remain concerning the perceptual capabilities of existing

B. Mutlu et al. (Eds.): ICSR 2011, LNAI 7072, pp. 11–20, 2011.

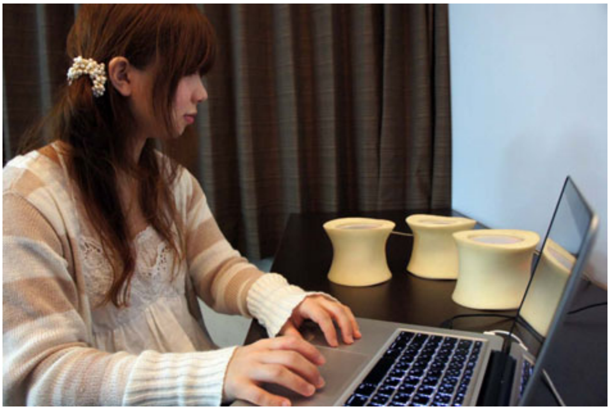


Fig. 1. The MAWARI-based interactive social interface

robots, a challenge to developing these areas still persists. One-to-one communication should consider the above essential factors with many response situations and frequently observe and maintain the above factors to exert effort (workload) in order to continue a conversation. In this sense, what will happen if we increase the involvement of parties to diminish their workload?

News reports on TV are mostly presented by a single reporter with the goal of disseminating details in a summarization format. Within this format it is difficult for one to participate in a discussion or conversation. However, some news channels report the news based on a discussion formation which involves a discourse that helps individuals acquire more detailed information about the news than is possible in the above scenario (a single reporter). But it is still difficult for outsiders to participate in a discussion in order to acquire their own information. Also, a certain amount of expertise is required to motivate individuals to be involved in the discussion and to align the discussion in the right direction. The above is construed as motivation to implement a multiparty-based conversation, which is most effective to utilize a robot as an information service medium.

Sakamoto and his colleagues [12] have been exploring the use of a robot as a broadcasting medium at a railway station. Using multiple robots (two Robovies) being fixed somewhere within the railway station, they conducted various experiments mainly concerned with a social and interactive setup in broadcasting. The robots mainly discuss information with passengers mainly listening to the conversation to gain necessary information (as a passive medium). Other contexts are defined as an interactive medium, which trace the passenger's availability in front of the robot through a laser range finder. If a passenger is presented in front of the robot, then the robot starts greeting it and the same conversation is presented in the context of the passive medium. This was almost similar to the news broadcasting, except that the robots participated in the conversation in real time. A comparison that is made with a second context is almost similar to the first context, except that the robot traces the passenger's position and greets him.

The advantage of a multiparty conversation is that each of those involved has its own personality and possesses a variety of knowledge [13]. The above factors are effective in producing a conversation to obtain necessary information when the robot is utilized as an information service medium or broadcasting medium. Mutlu and his colleagues [8] explore a conversation structure and the participants' roles and how these roles shift during a multiparty-conversation. The gaze direction is a crucial cue for participants to signify the status of role shifting [9][7]. In human-robot interactions, researchers have studied and explored the significance of communication channels (e.g., body gestures and eye gaze) in producing a conversation between participants. These studies have attempted to combine the turn-talking management skills of a robot for sufficient interaction capabilities to create a successful conversation. During a multiparty-conversation, the participants must dynamically and frequently observe and adapt to the conversation, and these factors are directed so that participants have a considerable workload (e.g., to trace the behaviors and to process the turn-taking management etc.) during the conversation [17].

In contrast, we propose a MAWARI-based (creatures) social interface as an interactive social medium to broadcast information (e.g. news, etc.). The MAWARI-based novel interface changes to either a passive or interactive social state in order reduce the workload of the conversation. The passive social interface is a state whereby the user (human) does not participate in the conversation, but creatures conduct the multiparty-discourse (broadcasting the information) and the user acquires information without participating. But the situation will switch to an interactive social interface when the user (human) expects to participate in the conversation to acquire either more interesting information or to align the conversation in the right direction. This paper describes the designing mechanism of creature and evaluates the workload of the user during a multiparty-conversation as in an interactive social medium.

2 Interface of MAWARI as an Interactive Social Medium

Goffman [4] introduced the concept of "footing," which depicts the role of participants in a conversation and also describes the concept of shifting roles in understanding the social interaction to realize the entailing of the conversation. In more than two party conversations (multiparty), we can define the four main roles of participants as that of speaker, address, side participants, and bystanders [8][16]. During a conversation, the participants' roles rotate. One of the participants plays the role of a speaker and one plays the role of an addresser. The side participant waits for the conversation, but the unaddressed person (at a particular moment) and the bystander do not contribute to the conversation [6].

The proposed interface of MAWARI is based on the above criteria in order to reduce the conversation workload due to the shifting roles of a user during a conversation (Figure 1). However, the MAWARI interface has two main attributes: one as a passive social interface and another as an interactive social interface. The state of the passive social interface is defined as when the user does not

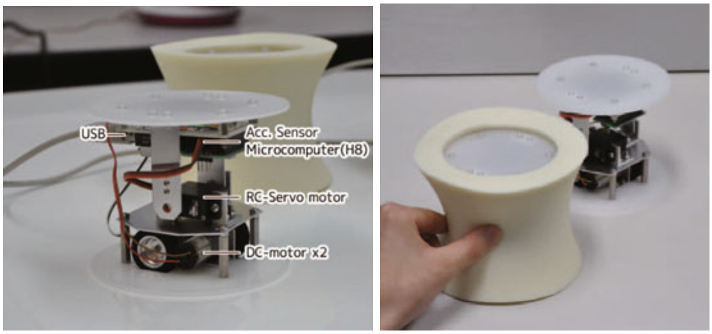


Fig. 2. MAWARI's external and internal structure with simple designing mechanism

participate in the conversation but listens and acknowledges the conversation without exerting any effort. In this context, the user acts as a bystander in the conversation and shifts the role to that of a side participant, address, or speaker when the interface becomes socially interactive. The user utilizes the MAWARI-based interface as an interactive social interface when attempting to gain more information or the conversation is aligned in the right direction.

The objective in altering the attribute of the interface is to reduce the conversation workload while changing the roles of the conversation. However, during the conversation the user does not need to answer or respond to the conversation. Instead, another MAWARI takes over the responsibilities (roles) to continue the conversation (i.e., as a passive social interface). However, there are times when the user has to participate in the conversation as described previously (i.e., as an interactive social interface). The performance of the interface should shift the interface state in order to reduce the workload of the user. Three creatures are required to continue a conversation in a productive way (i.e., user satisfies the contents of the information). Then, as long as the user's role remains that of a bystander (i.e., interface state remains passive) and the three creatures interact with the user when he participates in the conversation, the conversational roles change to the side of the participant, addressee, and speaker (interface state remains as an interactive social type). In the current study, we were interested in evaluating the workload of the user (in a conversation) when the interface of MAWARI is in a passive social mood. The current MAWARI setup is capable of broadcasting the news through a multiparty-discourse. But we believe that the MAWARI-based interface can be utilized in different applications as an interactive social medium.

3 Designing Concept of MAWARI

MAWARAI is designed with a minimal designing mechanism, contains a round-shaped body without eyes, and is capable of moving by wheel and changing its attractive body gestures during a conversation. The motivation behind the novel design is to reduce the number of modalities involved in a conversation

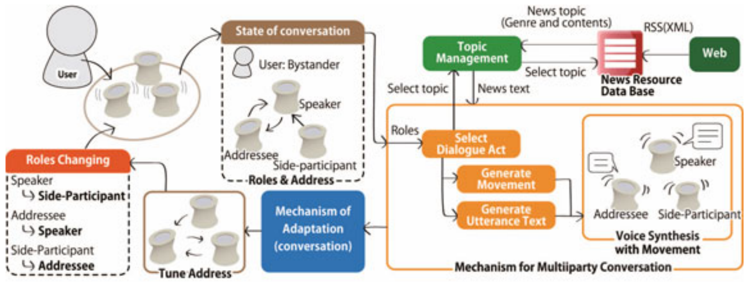


Fig. 3. The architect of MAWARI for generating a conversation, voice, and gestures for interact with attractive social behaviors

in order to reduce the workload of participants (i.e., no need to follow eye-gaze, facial expression, life movement, etc.). Its external material is made of soft touch urethane foam (Figure 2(right)), and it is mainly designed to interact on the table. Creatures are capable of nodding the upper part of its body, and its whole body can also move back-and-forth or turn. These attractive social cues can be used during the conversation to intensify the social rapport between the user and creatures.

The microphone and speaker connect to the main computer, and the inside of the computer processes voice recognition, voice synthesis, and generation of its conversation, which is obtained through RSS in real time. MAWARI is connected by a main computer via a USB cable, and behaves based on the information of the control module from the main computer. Creature has a microcomputer (H8), acc sensor, RC-servomotor and two DC-motors. These motors are controlled by the PWM output of the microcomputer through the motor driver IC, and the DC-motors are arranged on the basement. The RC-servomotor is arranged in the middle of body, which allows the creature to have a nodding behavior. A three-dimensional acceleration sensor is mounted to detect the creature's postural stability (Figure 2(left)).

Figure 3 depicts the architect of MAWARI for establishing the multi-party conversation in a passive social interface. The MAWARI-based interface extracts the topics (news) from RSS in real-time to continue the conversation. Each of the MAWARIs generate their conversations according to the text of the news and dialogue acts, which is referred to in DAMSL[1]. The conversational roles are determined according to the state of the conversation (history), the characteristics of MAWARI (according to the news topic), and topic management. MAWARI behavior is generated according to the rules of the dialogue acts and the topic of news. Multi-party conversation is continued by updating the state of the conversation and topic of the news, which is updated by the mechanism of adaptation. We used the voice recognition open source engine Julius and the voice synthesis

[1] DAMSL: http://www.cs.rochester.edu/research/cisd/resources/damsl/RevisedManual/

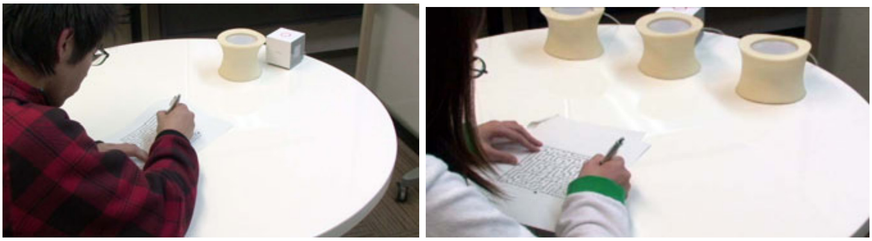


Fig. 4. The user solves a maze game while listening and understanding the contents of the news being broadcast during the experiment

engine Wizard Voice (ATR-Promotions) and web service (RSS) by XML format to gather information from the Web. MAWARIs use extracted topics (genre and text) from a database to generate the contents of the conversation.

4 Empirical Evaluation

The dual task of the experiment was for the user to solve a maze game, as well as listen to and understand the contents of the news being simultaneously broadcasted from the creatures. The goal of the experiment was to compare the workload of the user when a single creature with a passive social interface is used versus a multiple creatures (three) with a passive social interface (Figure 4). Also, we attempt to determine how much the MAWARIs contributed or facilitated the performance of multi-tasks (dual tasks) for the user.

The experimental method used in this work was as follows. The user participated in two trials (scenario A and scenario B) with the MAWARI interface. Six news sources were used in the conversation. In the initial trial (scenario A), the user was required to solve a maze game while listening and understanding the contents of the news being broadcast from a single MAWARI. During the first trial (interaction with a single creature) the MAWARI broadcasted the news source in combination with attractive social cues (bending, twist, etc.) through body gestures based on attributes of the speech. In the second trial (scenario B), the user was required to follow the same procedure as above, but three creatures participated in the conversation. In the second trial (scenario B), the MAWARIs randomly distributed six news sources to broadcast the news among the MAWARIs (multiparty-conversation) and performed nodding behaviors, which provided encouragement to participants, enabling them to adapt and adjust to the conversation. The content of the news was almost similar in both trials (scenario A and scenario B) in the conversation. Half of the users participated first in trial 1 (scenario A) and then in trial 2 (scenario B), while the other half participated first in trial 2 and then in trial 1. Such a strategy is helpful to obtain a counter-balance of the data, thereby reducing the effect of the sequence of trials on the results.

After each trial was finished, the user was required to answer six questions to evaluate the workload of the user from NASA-TLX (see Table 1). We used

Table 1. Table shows the dimensions of the NASA-TLX based on our context and the questionnaire of the user

Dimensions	Endpoints	Questions	Descriptions
Mental Demand	Low/High	Can you remember the contain of the news?	How much mental and perceptual activity was required in thinking, deciding, remembering, searching, and etc?
Physical Demand	Low/High	Method of presenting the information was boring or tired?	How much physical activity was required turning, observing, controlling, etc.?
Temporal Demand	Low/High	Did you feel time pressure during the conversation?	How much time pressure did you feel during the conversation; was there a pressure to participate?
Performances	Good/Poor	Did you accomplish your tasks during the interaction?	How successful do you think you were in accomplishing the goals?
Effort	Low/High	Did you put effort into the conversation to obtain information?	How did you have to work to accomplish your level of performances?
Frustration	Low/High	Did you feel frustrate during the conversation?	How insecure, discouraged, irritate, stress and complacent did you feel during the tasks?

the strategy of dual task context with NASA-TLX (NASA-Task Load Index) [5] to evaluate the workload of user. For each question, a rating scale (1-10) was used to rate the questions, and the user was required to answer two questions to evaluate their feelings and the social characteristics of the MAWARI-based interface. A total of sixteen students (aged between 19 and 22 years) participated in the experiment.

4.1 NASA-Task Load Index (NASA-TLX)

NASA-TLX is an index containing a multi-dimensional rating system consisting of six dimensions: mental demand, physical demand, temporal demand, performances (own), effort, and frustration. The above weighted average of rating is useful to evaluate the overall workload of a user. A questionnaire was designed to represent the above dimensions as describe in Table 1. The degree of the above six factors contributed to representing the workload of the user which is represented from the rates of user perspectives. These rates were determined by the user's response to pair-wise comparisons among the six factors. After assessing the magnitude of each of these six factors on a scale, each individual performed pair-wise comparisons between six factors to determine the higher source of workload factor for each fair.

4.2 Results

The dimension of the NASA-TLX is defined as follows (Table 1) in our context, which was slightly changed to the original [11][1][2]. We estimated the NASA-TLX values for each of the dimensions based on the subjective rating for scenario A and scenario B. A pairwise t-test was applied to determine if the difference between scenario A and scenario B was significant or not. The results, shown in Table 2, revealed that the difficulty of memorization (mental demand) in the

Table 2. Tables show the difference in mean values for the NASA-TLX values in scenario A and scenario B

Dimentions	Mean value (standard error)		pairwise t-test t(d.f.) = t-value	Results
	Scenario A	Scenario B	$p < 0.05$ (significant)	
Mental Demand	4.69 (0.60)	5.06 (0.48)	t(15) = 0.78 $p > 0.05$	Not Significant
Physical Demand	5.50 (0.68)	4.06 (0.58)	t(15) = 1.88 $p < 0.05$	Significant
Temporal Demand	5.31 (0.63)	4.19 (0.65)	t(15) = 1.50 $p > 0.05$	Not Significant
Performances	7.19 (0.58)	5.50 (0.50)	t(15) = 2.39 $p < 0.05$	Significant
Efford	6.50 (0.47)	7.00 (0.40)	t(15) = 1.12 $p > 0.05$	Not Significant
Frustration	4.75 (0.73)	3.75 (0.66)	t(15) = 1.69 $p > 0.05$	Not Significant

two scenarios was insignificant, indicating that the user did not have any problems in remembering the contents of the news in both situations. In the physical demand dimension, the NASA-TLX value showed a significant difference, with the mean value of scenario B being low. This indicates that when a user faced a one-to-one interaction (scenario A), the user had to perform several physical activities (motivated to ask questions, paying attention to the robot's behaviors, and turning). The user had the expectation that other MAWARIs would participate in the conversation, meaning the user was not required to exert physical energy into the scenario B (i.e., the rating was low). The other dimensions of time pressure, effort, and frustration had a nonsignificant effect in either of the scenarios. However, the dimension "performance" showed a significant difference in rating, indicating that scenario B received a higher rating in performance than did scenario A. The overall raw score (R-TLX: t(15 d.f)= 1.84, $P < 0.05$) yielded a significant difference for scenario A and scenario B, and the adaptive weighted workload (AWWL: t(15 d.f.)= 1.90, $P < 0.05$) also showed a significant difference for scenario A and scenario B. These two synthesis values showed that condition B (scenario for a multiparty conversation with three MAWARIS) had a lower mental workload in the interaction or conversation.

5 Discussion and Conclusion

Overall, the results showed that in both scenarios there were no significant differences in the dimensions of mental demand (memorizing difficulty), time pressure, effort, and frustration. The dimension of memorization difficulty indicates that the MAWARI-based passive social interface did not contain the workload of remembering, thinking, and searching, which affected both scenarios. The results of both the time pressure and effort dimensions indicate that the interface had no significant differences. However, the interface was capable of establishing a well-disposed social rapport between the users and creatures, because users were not pressured to be concerned about the time coordination of the interaction and therefore experienced no frustration during the conversation. The one-to-one conversation required the exertion of more effort in regards to physical demands (observing eye gaze behavior, turn-taking, etc.) to create a better and more estimable conversation with the partner.

Similar results were found when a user interacted with a single MAWARI, because the user reflexively considered what the physical demands were in order to establish interaction. But with three MAWARIS (multiparty-conversation) the user did not place any physical demands on the conversation, because the user expected the other MAWARIs to switch roles in order to continue the conversation. These results suggest that the user workload of the conversation was reduced by the multiparty-conversation with passive social interface. The experimental procedure was mainly designed to evaluate the mental workload of the user on the MAWARI-based passive social interface on a dual-task. The users' ratings indicated that there were significant differences in performance in the dual task within both scenarios.

In general, scenario B had a higher performance than scenario A, as the user was able to perform multi-tasks when the interface was in the passive social mode with a multiparty-discourse. The overall rating of the raw overall score (R-TLX) adaptive weighted workload (AWWL) indicated that both scenarios had significant differences in the mental workload, and that there was less mental workload in the context of the multiparty-conversation in comparison to the one-to-one conversation. We believe that the dimensions of "physical demand" and "performance" substantially affected the values of R-TLX and AWWL resulting in a significant difference on the mental workload in both scenarios. Furthermore, the overall results indicate that users critically considered the physical demands (eye gaze behaviors, turn-taking, thinking, etc.), and performance (of multi-tasks) on the mental workload. The results also indicate that our passive social interface creature design (following the minimalism designing concept) helped to facilitate the reduction of the user workload. It also provided a physical relief in performing multi-tasks while having multiparty-discourse.

Finally, users answered a questionnaire (two questions) about whether the interface was friendly or not, as well as whether the conversation was easily accepted or not. The results of the first question showed there were significant differences between the two scenarios, with condition B receiving a higher rating from users. Therefore, we believe that the users and three MAWARAIs together established an effective social interaction with prolific conversation as a social information medium. The results of the second question revealed there were significant differences in user ratings, indicating that the MAWARI-based interface was able to perform within conversations in a friendly and socially acceptable way.

Acknowledgments. This research has been supported by both Grant-in-Aid for scientific research of KIBAN-B (21300083) and Grant-in-Aid for JSPS research fellow (10JO6701) from the Japan Society for the Promotion of science (JSPS).

References

1. Bortolussi, M.R., Kantowitz, B.H., Hart, S.G.: Measuring pilot workload in a motion base trainer: A comparison of four techniques. Applied Ergonomics 17(4), 278–283 (1986)

2. Byers, J.C., Bittner, A.C., Hill, S.G., Zaklad, A.L., Christ, R.E.: Workload assessment of a remotely piloted vehicle (rpv) system. In: Proceedings of the Human Factors Society Thirty-Second Annual Meeting, pp. 1145–1149 (1988)
3. Duncan, S.: Some signals and rules for taking speaking turns in conversations. Journal of Personality and Social Psychology 23, 283–292 (1972)
4. Goffman, E.: Footing. Semiotica 25, 1–29 (1979)
5. Hart, S.G., Stavenland, L.E.: Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: Hancock, P.A., Meshkati, N. (eds.) Human Mental Workload, ch. 7, pp. 139–183. Elsevier (1988)
6. Herbert, T.B.C., Clark, H.: Hearers and speech acts. Language 58, 332–373 (1982)
7. Kendon, A.: Some functions of gaze-direction in social interaction. Acta Psychologica 26, 22–63 (1967)
8. Mutlu, B., Shiwa, T., Kanda, T., Ishiguro, H., Hagita, N.: Footing in human-robot conversations: how robots might shape participant roles using gaze cues. In: HRI 2009: Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction, pp. 61–68. ACM, New York (2009)
9. Nakano, Y.I., Ishii, R.: Estimating user's engagement from eye-gaze behaviors in human-agent conversations. In: IUI 2010: Proceeding of the 14th International Conference on Intelligent User Interfaces, pp. 139–148. ACM, New York (2010)
10. Okada, M., Sakamoto, S., Suzuki, N.: Muu: Artificial creatures as an embodied interface. In: 27th International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH 2000), The Emerging Technologies: Point of Departure (2000)
11. Rehmann, A.J.: Handbook of human performance measures and crew requirements for flightdeck research
12. Sakamoto, D., Hayashi, K., Kanda, T., Shiomi, M., Koizumi, S., Ishiguro, H., Ogasawara, T., Hagita, N.: Humanoid robots as a broadcasting communication medium in open public spaces. I. J. Social Robotics 1(2), 157–169 (2009)
13. Suzuki, N., Takeuchi, Y., Ishii, K., Okada, M.: Talking eye: Autonomous creatures for augmented chatting. Robotics and Autonomous Systems 31, 171–184 (2000)
14. Thrun, S., Bennewitz, M., Burgard, W., Cremers, A.B., Dellaert, F., Fox, D., Hähnel, D., Rosenberg, C.R., Roy, N., Schulte, J., Schulz, D.: Minerva: A second-generation museum tour-guide robot. In: ICRA, pp. 1999–2005 (1999)
15. Trafton, J.G., Bugajska, M.D., Ratwani, R.M., Fransen, B.R.: Integrating vision and audition within a cognitive architecture to track conversations. In: Proceedings of the Human Robot Interaction, HRI (2008)
16. Vertegaal, R., Slagter, R., et al.: Eye gaze patterns in conversations: There is more to conversational agents than meets the eyes, pp. 301–308. ACM Press (2001)
17. Yoshikawa, Y., Shinozawa, K., Ishiguro, H., Hagita, N., Miyamoto, T.: Responsive robot gaze to interaction partner. In: Proceedings of Robotics: Science and Systems (2006)

Design of Robust Robotic Proxemic Behaviour

Elena Torta*, Raymond H. Cuijpers, James F. Juola, and David van der Pol

Eindhoven University of Technology,
Den Dolech 16, 5600MB Eindhoven, NL
{e.torta,r.h.cuijpers,j.juola,d.v.d.pol}@tue.nl

Abstract. Personal robots that share the same space with humans need to be socially acceptable and effective as they interact with people. In this paper we focus our attention on the definition of a behaviour-based robotic architecture that, (1) allows the robot to navigate safely in a cluttered and dynamically changing domestic environment and (2) encodes embodied non-verbal interactions: the robot respects the user's personal space by choosing the appropriate distance and direction of approach. The model of the personal space is derived from a human-robot psychophysical study and it is described in a convenient mathematical form. The robot's target location is dynamically inferred through the solution of a Bayesian filtering problem. The validation of the overall behavioural architecture shows that the robot is able to exhibit appropriate proxemic behaviour.

Keywords: Robotic Navigation, Particle Filter, Proxemic.

1 Introduction

Personal robots that share with humans the same environment need to be socially acceptable and effective when interacting with the environment as well as the users [19]. Embodied non-verbal interactions, such as approach, touch, and avoidance behaviours, are fundamental to regulating human-human social interactions, and they are likely to play a similar role in human-robot interaction [4]. Extensive research has been conducted in the social science domain to model people's proxemic behaviour and to assess how their mutual position influences the quality of interaction, see Hall [5] and recently Lambert [8]. General results show that relative behaviour and positioning is influenced by the task to be accomplished and the degree of familiarity that exists between the different actors [15],[14] and [10]. The physical embodiment of robots makes it likely that they will have to follow societal norms in establishing their physical and psychological distancing with people [19],[10]. Therefore the trajectories of a socially assistive robot, that is supposed to live in a domestic environment while sharing the same space with a human being, represent a non-verbal communication cue that influences the quality of interaction. Producing efficient trajectories is not enough.

* The author gratefully aknowledge the work done by James Loma, Joey and Richie Baten and Frank Basten.

B. Mutlu et al. (Eds.): ICSR 2011, LNAI 7072, pp. 21–30, 2011.

Trajectories and end poses (robot's final location and orientation) need to be effective in the sense that the robot modifies its proxemic behaviour according to the task to be accomplished, to the environment and to user's expectations [6],[7],[12]. We present interdisciplinary work that addresses navigation issues related to the robot's proxemic behaviour presenting methodologies and results that lie in the psychological and technological domain.

2 Navigation Framework

Navigation is a necessary capability for mobile robots to move around in their surroundings. In this section we briefly present a navigation system that relies on the attractor dynamic approach to behaviour-based robotics, see [3] and [11] for the theoretical background and [1] and [9] for technical applications. A behaviour can be described by means of a behavioural variable that, in our work, is chosen to be the robot's heading direction, $\phi(t)$, and by the temporal evolution of it. The evolution is controlled by a non-linear dynamical equation that can be generally expressed as:

$$\frac{d\phi(t)}{dt} = \sum_{i=1}^{m} w_i f_i(\phi(t)) + d, \tag{1}$$

where m represents the number of behaviours that are needed for the robot to accomplish its task. The term $f_i(\phi(t))$ represents the force produced by the i^{th} behaviour and w_i represents its weight. The term d represents a stochastic term that is added to guarantee escape from repellers generated by bifurcation in the vector field [9]. We identify three basic behaviours whose coordination brings the robot from a generic location in the environment to a target location. The process of reaching a target point is represented by an attractor dynamic:

$$f_1(t) = -\sin(\phi(t) - \psi_{tar}(t)), \tag{2}$$

where the term $(\phi(t) - \psi_{tar}(t))$ accounts for the angular location of the target with respect to the robot at time t. The attractor dynamic acts to decrease the difference between $\phi(t)$ and ψ_{tar}. The graphical representation of those angles is visible in Fig. 1. The ability to obtain collision-free trajectories is encoded by a repulsive dynamic whose mathematical expression is given by:

$$f_2 = \exp\left(-\frac{1}{\beta_2}\left(d_{obs}(t) - R\right)\right)\left(\phi(t) - \psi_{obs}(t)\right)\exp\left(-\frac{\left(\phi(t) - \psi_{obs}(t)\right)^2}{2\sigma_{obs}^2}\right) \tag{3}$$

It generates a force which is dependent exponentially on the detected distance, $d_{obs}(t)$ between the robot and the obstacle through the term $\exp\left(-\frac{1}{\beta_2}\left(d_{obs}(t) - R\right)\right)$ and on the angular location of the obstacle with respect to the robot thorough the term $\exp\left(-\frac{(\phi(t)-\psi_{obs}(t))^2}{2\sigma_{obs}^2}\right)$.The coefficients β_2 and σ_{obs} determine the range at which the repulsion strength acts. In order to obtain collision-free trajectories, the repulsive weight w_2 is greater than w_1. The graphical visualization of $\psi_{obs}(t)$ is reported in Fig. 1. The alignment behaviour that

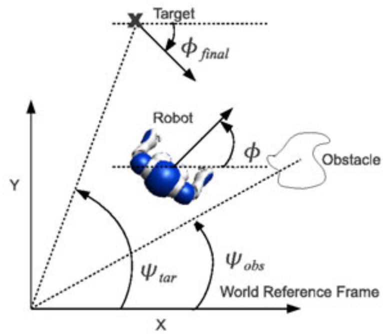


Fig. 1. Graphical representation of the robot, the target and an obstacle. The angles $\phi(t)$, ψ_{tar}, ϕ_{final} and ψ_{obs} are reported with respect to a global reference frame. The desired robot's final orientation is represented as target orientation.

acts on $\phi(t)$ to reduce the angular distance between the robot's actual orientation and the desired final orientation, ϕ_{final}, is also modelled by an attractor dynamic:

$$f_3(t) = -\exp\left(-d_t(t)\right)\sin(\phi(t) - \phi_{final}), \tag{4}$$

where the term $(\phi(t)-\phi_{final})$ represents the difference between the robot's orientation at time t and the requested final orientation. The graphical representation of ϕ_{final} is visible in Fig. 1. The attractor dynamic of the alignment behaviour is undesirable when the robot is far from the target point. This consideration is modelled by the term $\exp\left(-d_t(t)\right)$ of Eq. (4) that decreases the attractor dynamic of the behaviour exponentially with the distance, $d_t(t)$, between the robot and the target point. The robot's forward should be low when the robot is near obstacles or when the angular distance between the robot and the target is large. These considerations are taken into account in the following expression of the forward velocity:

$$v_{for} = \min\left(\exp\left(-\frac{1}{d_{obs}(t)+\varepsilon}\right), \exp\left(-|\phi(t) - \phi_{final}(t)|\right)\right)\cdot v_{max}, \tag{5}$$

where ε is a term added to avoid a division by zero and v_{max} is the maximum robot's forward speed. The forward velocity v_{for} decreases exponentially with the decrease of the distance, $d_{obs}(t)$, between the robot and the obstacle or with the angular distance between the robot's actual heading angle and the desired target final orientation through the term $\exp -(|\phi(t)-\phi_{final}(t)|)$. A similar navigation architecture has been adopted in several studies [3] ,[13], [2],[1] and [9]. In all this work, emphasis has been given to the precise definition of the behaviours to allow the robot to generate collision-free trajectories in a cluttered environment. Little attention has been given to the description of the target itself. But when the robot is supposed to address a human being, the definition of such a location is dependent on what the human expects from the robot in terms of distance and direction of approach. The selection of an appropriate location of approach determines how the user will perceive the overall robot's proxemic behaviour, thus it influences the quality of interaction. It is therefore necessary to derive quantitative models of what users expect in terms of distance and direction of approach from an autonomous robot that can be easily embedded in already existing navigation frameworks with the purpose of dynamically defining the robot's target point.

3 User Study

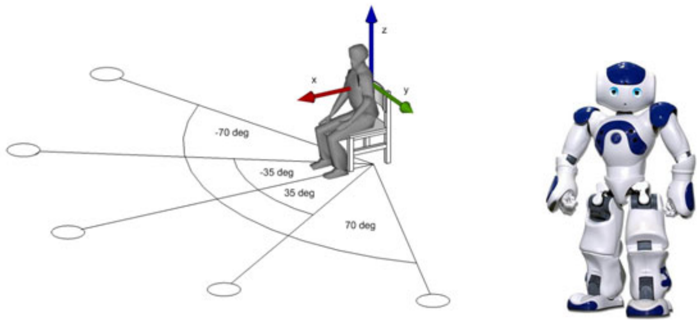
In this section we determine a model of the Personal Space (PS) by means of a psycho-physical experiment in a scenario in which the robot NAO (see Fig. 2(b)) by Aldebaran Robotics, approaches a person for conveying information. The model's expression is then included into the navigation framework.

3.1 Methodology

The psycho-physical study was carried out in a laboratory that approximates a living room. The total number of participant was 10, 8 males and 2 females. The mean age of subjects was 24 years (range: 16-51). The subjects had a technical background but were not familiar with robotics research. The subject was seated in a chair and had free space in all directions, see Fig. 2(a). The robot was placed on the starting point by an operator and was driven under direct remote control to approach the subjects following a straight line. The five robot's starting points were defined with respect to the user reference frame, a right-handed reference frame with the y-axis aligned to the subject's shoulders and the x-axis pointing straight ahead. The starting points are defined as $\chi = \{(4m, -70°), (4m, -35°), (4m, 0°), (4m, 35°), (4m, 70°)\}$.Subjects were instructed to keep their back on the chair but could move their head freely. The operator was seated in the control room next to the living room. Subjects had to stop the robot by pressing a button. The desired moment to stop the robot was task dependent and three tasks were assigned to the subjects:

- Task A (Optimal): the robot is stopped when it arrives at the best distance for having a comfortable communication.
- Task B (Close): the robot is stopped when it arrives at the closest distance for having a comfortable communication.
- Task C (Far): The robot is stopped as soon as it arrives at a distance which is close enough for having a comfortable communication.

When the robot stopped it said "Hello" to the subjects. After every trial, the distance between the robot and the person was recorded via a ceiling-mounted camera. Subjects were also requested to evaluate the direction at which the robot approached them by answering a questionnaire. The latter consisted of a question with a 5-point Likert scale that ranged from Very Good to Very Bad. Each task (A,B,C) had to be performed three times by each subject for every direction therefore the user study used a 3 (task) x 5 (direction) design, with three replications in each factorial combination yielding a total of 45 trials per participant.

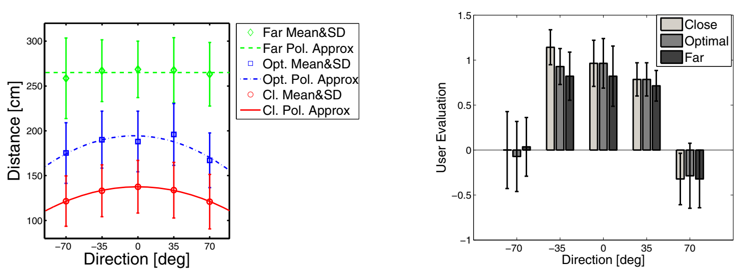


(a) Set Up (b) Nao

Fig. 2. (a) Graphical representation of the experiment set-up. The user was seated on a chair and was approached by the robot NAO from five different directions. The different starting points, the straight trajectories and the user reference frame are also reported. (b) The robot Nao.

3.2 Results

The data collected allowed the derivation of quantitative relationships to describe the shape of the PS in an approaching scenario. The relationship between distance and direction of approach for task A (optimal distance) and task B (closest distance) can be approximated via a second order polynomial, while for task C (furthest distance), the relationship is constant. The goodness of approximation for the three tasks has been verified with an ANOVA test. The results are summarized in Table 1 and a graphical representation is visible in Figure 3(a).



(a) (b)

Fig. 3. (a) Association between encounter direction and optimal distance in a conveying information scenario. (b) Bar chart representation of the mean value of the subjects evaluation per approaching direction per task with associated 95% confidence interval. Negative values are representative of negative judgements whereas positive values represent positive judgments.

The user study also allowed to evaluate how subjects perceived the direction of approach. To clearly identify the difference between comfortable directions and uncomfortable directions, negative numbers were associated with negative

Table 1. Summary of the parameters of the approximating curves that relate distance and direction of approach for the three tasks of the user study. The parameters for evaluating the goodness of approximation are also reported.

Pol. ID	Pol. order	ANOVA	Coefficient
ρ_A	2	F= 5.6, p=0.004	$\zeta_{2_A} = -14.9$, $\zeta_{1_A} = 0$, $\zeta_{0_A} = 194.4$
ρ_B	2	F=3.8, p=0.025	$\zeta_{2_B} = -10.9$, $\zeta_{1_B} = 0$, $\zeta_{0_B} = 137.6$
ρ_C	0	—	$\zeta_{0_C} = 270.0$

items of the Likert scale and *vice versa*. The values are: very bad (-2), bad (-1), neutral (0), good (1) and very good (2). The mean value of users' rating was computed per direction of approach per task and the results are displayed in Figure 3(b). The user evaluation of the approaching direction shows that there is a significant difference between the central directions of approach ($\pm 35°, 0°$) and the furthest directions ($\pm 70°$) and that farthermost directions are perceived as uncomfortable while the frontal directions are in general perceived as comfortable. Unlike the results of the user study conducted by Dautenhahn et al. [4] we did not find a significant difference between approaching from the right side or from the left side. It is possible to derive a mathematical relation between the direction of approach and the preference associated to that direction. At each item of the Likert scale we now associate positive numbers, between 0 (very bad) and 1 (very good). The relationship between direction of approach and the user's preference for the direction is approximated via a second-order polynomial. The goodness of approximation for the three tasks has been verified with an ANOVA test and the results are summarized in Table 2.

Table 2. Summary of the parameters of the approximating curves that relate directions of approach to the user's evaluation of them for the three tasks of the user study. The parameters for evaluating the goodness of approximation are also reported.

Pol. ID	Pol. order	ANOVA	Coefficient
ν_A	2	F=31.0, $p < 0.001$	$\kappa_{2_A} = -0.21$, $\kappa_{1_A} = 0$, $\kappa_{0_A} = 0.77$
ν_B	2	F=7.0, $p = 0.001$	$\kappa_{2_B} = -0.09$, $\kappa_{1_B} = -0.05$, $\kappa_{0_B} = 0.70$
ν_C	2	F=24.5, $p < 0.001$	$\kappa_{2_C} = -0.18$, $\kappa_{1_A} = 0$, $\kappa_{0_A} = 0.74$

3.3 Personal Space and Region of Approach

The data reported in Table 1 and 2 allow the definition of a region of approach centred at the optimal distance, Task A, that lies within the boundaries defined by the results for Task B (closest distance) and Task C (farthest distance). The proposed expression of the Region of Approach (RA) is given by:

$$RA = \{\mathbf{x} \in \mathbb{R}^2 : l(\mathbf{x}) \geq 0.1\} \tag{6}$$

where

$$\mathbf{x} = (\rho, \theta)$$

$$l(\mathbf{x}) = l(\rho, \theta) = \nu_A(\theta) \exp\left(-\frac{(\rho - \rho_A(\theta))^2}{\sigma_{\rho_A}^2}\right), \tag{7}$$

where ρ represents the distance and θ represents the angular distance from the user. Expression (6) allows to define a region of space that associates at each point a value $l(\rho, \theta)$ that is dependent on the distance and direction of approach and that expresses how comfortable that location is for the user in a scenario in which the user sees the robot approaching for conveying information. The value of acceptability is directly dependent on the direction of approach through the evaluation of the term $\nu_A(\theta)$, see Table 2, and on the approaching distance through the evaluation of the term $\rho_A(\theta)$, see Table 1. The standard deviation σ_{ρ_A} represents the one associated to the recorded distances of task A (optimal) and it is equal to 34.13 cm. A graphical visualization of Eq.(6) is visible in Fig. 4.

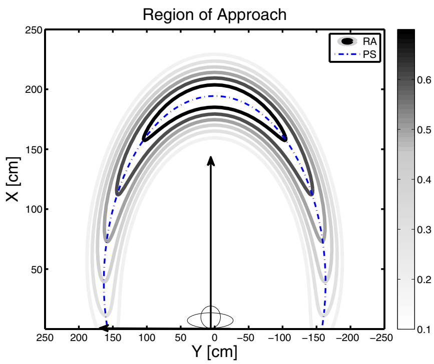


Fig. 4. Contour graph of the Region of Approach (RA). A value $l(\rho, \theta)$ (see Eq. 6) is assigned to each point proximate to the user. The value is representative of the subjects' rating of that point in an approaching scenario.

3.4 Target Representation and Navigation

The region of approach defined in Eq.(6) can be regarded as a measurement of how the user evaluates the space around from a proxemic HRI perspective and can be introduced in a framework for the dynamic selection of the target. Indeed we notice that the region of approach introduced in expression (6) is suitable for Bayesian inference with the objective of deriving the best robot position, proximate to the user, given the knowledge of user's preferences and the robot's perception of the environment. Such observation allows to apply the well established particle filter technique for dynamically allocating the target point. An overview on Bayesian filtering techniques for robotic applications can be found in [16]. More details on the framework applied to proxemic HRI can be found in authors previous work [17] and [18]. The state to be estimated is the best robot end-pose in the user proximity and the algorithm applied to the recursive Bayesian filtering problem is a particle filter. The weight expression of the particle filter is given by :

$$w^i = \frac{1}{d(x_t^i, x_r)} RA(x_t^i) \tag{8}$$

where $d(x_t^i, x_r)$ represents the distance between the robot and the particle x_t^i at time t and $RA(x_t^i)$ represents expression (6) evaluated at the particle location. The presence of the term $d(x_t^i, x_r)$ in the weighting expression (8) allows the model to cope with the presence of obstructions since the the repulsive force, described in Eq.(3), drives the robot away from obstructed locations, and thus the weights associate to particles in those locations decrease. The term $RA(x_t^i)$ allows constant consideration of the user's expectations in terms of robot's proxemic behaviour. The product of the terms in Eq.(8) indicates that a particle is positioned in a desirable location if it simultaneously represents a feasible point, thus a low value of $d(x_t^i, x_r)$, and a desirable point in terms of user's preferences, thus a high value of $RA(x_t^i)$). Indeed the conceptual novelty of the particle filter presented here is the fusion of contextual cues derived from the robot's perception with knowledge derived from psychological experiments.

4 Validation

The effectiveness of the navigation algorithm enhanced with the model of the user's personal space and the dynamic target selection was tested during a technical trial. The validation set-up was the same as the experiment set-up. The robot approached a seated person for conveying information. The robot autonomously determines its target point and therefore its trajectory. The position of the user with respect to the robot is given externally by a ceiling-mounted camera. A screen-shot from the ceiling-mounted camera during the validation trial can be seen in Figure 5. The recorded trajectories of the robot and the target, expressed with respect to the user's reference frame are visible in Fig. 6. The target location dynamically changes over time due to the dynamic inference process and its location is influenced by the RA as well as the robot's trajectory. The robustness of the algorithm was also tested. The robot was requested to approach a person starting from a frontal position. An obstacle is placed on the initial target location. As soon as the robot modifies its trajectory, due to the presence of the

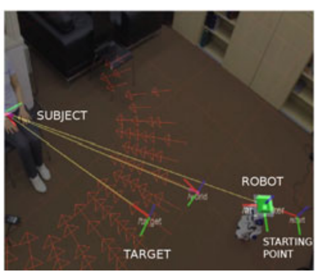
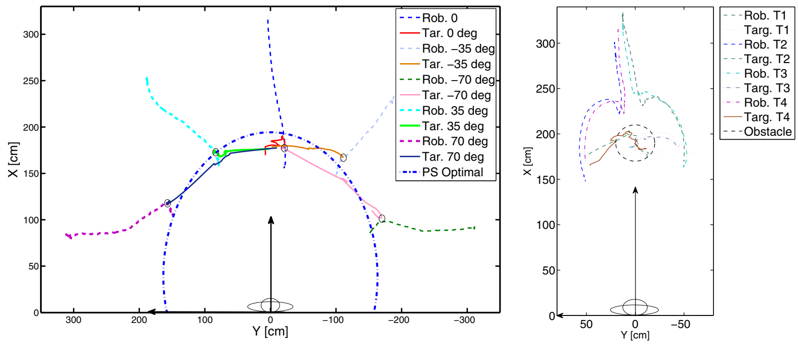


Fig. 5. A screen-shot from the validation trial. The particles are represented by the red arrows while the inferred target location is identified by the reference frame labelled as Target.



(a) Free Space (b) Obstacle

Fig. 6. Technical validation of the navigation architecture enhanced with the PS model. (a) Robot's trajectories and dynamic evolution of the target location. The point where the robot reaches the target is shown with a circle. (b) Robot's and target's trajectories in the presence of an obstacle.

detected obstacle, the target location changes allowing the robot to cope with the presence of the obstacle approaching the person from a comfortable location, see Figure 6(b).

5 Discussion and Conclusion

We presented a behaviour-based navigation architecture that allows a mobile robot to exhibit appropriate proxemic behaviour when interacting with a person. To do so, we derived a new model of the user's personal space in an approaching scenario that takes into account the relationship between direction and distance of approach and the user's evaluation of the approaching direction. We than applied a Bayesian filtering technique to dynamically infer the robot's target location for the navigation algorithm. The use of Bayesian inference with a mixture of psychological and contextual cues constitutes a novelty within the robotic navigation community and future development will integrate other contextual cues such as the user's head pose or physiological signals such as the galvanic skin conductivity. We validated the effectiveness of the navigation algorithm via real world trials. The trials have shown the feasibility and robustness of the overall navigation architecture. At the moment of writing the authors are running user studies to address how the inclusion of the personal space model in the navigation architecture improves the user's rating of the robot's proxemic behaviour and preliminary results are encouraging. The parameters of the personal space model are valid for the robot NAO, but their validity for other robots need to be further investigated. Future development may derive a standard parametric model of the personal space in a HRI scenario considering, as parameters, the robot's height, its appearance and the purpose of the interaction. We also considered that during the interaction the user remains seated on the chair therefore, for inference purposes, its state remains constant. Future development will address multiple user's states.

Acknowledgments. The research leading to these results is part of the KSERA project (http://www.ksera-project.eu) and has received funding from the European Commission under the 7th Framework Programme (FP7) for Research and Technological Development under grant agreement n2010-248085.

References

1. Althaus, P., Christensen, H.I.: Behavior coordination in structured environments. Advanced Robotics 17(7), 657–674 (2003)
2. Althaus, P., Ishiguro, H., Kanda, T., Miyashita, T., Christensen, H.I.: Navigation for human-robot interaction tasks. In: Proceedings of ICRA 2004, pp. 657–674 (2004)
3. Bicho, E.: Dynamic approach to Behavior-Based Robotics. PhD thesis, University of Minho (1999)
4. Dautenhahn, K., Walters, M., Woods, S., Koay, K.L., Nehaniv, C.L., Sisbot, A., Alami, R., Siméon, T.: How I serve you? In: Proceedings of HRI 2006, pp. 172–179 (2006)
5. Hall, E.T.: The Hidden Dimension. Doubleday, New York (1966)
6. Huettenrauch, H., Eklundh, K., Green, A., Topp, E.: Investigating Spatial Relationships in Human-Robot Interaction. In: IEEE/RSJ IROS 2006 (2006)
7. Kirby, R., Simmons, R., Forlizzi, J.: COMPANION: A Constraint-Optimizing Method for Person-Acceptable Navigation. In: RO-MAN 2009 (2009)
8. Lambert, D.: Body Language. Harper Collins, London (2004)
9. Monteiro, S., Bicho, E.: Attractor dynamics approach to formation control: theory and application. Autonomous Robots 29(3-4), 331–355 (2010)
10. Mumm, J., Mutlu, B.: Human-Robot Proxemics: Physical and Psychological Distancing in Human-Robot Interaction. In: Proceedings of HRI 2011 (2011)
11. Schöner, G., Dose, M., Engels, C.: Autonomous Systems Dynamics of behavior: theory and applications for autonomous robot architectures. Elsevier, Robotics and Autonomous Systems 16, 213–245 (1995)
12. Sisbot, E.A., Marin-Urias, L.F., Broquère, X., Sidobre, D., Alami, R.: Synthesizing Robot Motions Adapted to Human Presence. International Journal of Social Robotics (2010)
13. Soares, R., Bicho, E., Machado, T., Erlhagen, W.: Object transportation by multiple mobile robots controlled by attractor dynamics: theory and implementation. In: Proceedings of 2007 IEEE/RSJ IROS, pp. 937–944 (2007)
14. Syrdal, D.S., Koay, K.L., Walters, M.L., Dautenhahn, K.: A personalized robot companion? - The role of individual differences on spatial preferences in HRI scenarios. In: Proceedings of RO-MAN 2007, pp. 1143–1148 (2007)
15. Tapus, A., Mataric, M.J.: Socially Assistive Robots: The Link between Personality, Empathy, Physiological Signals, and Task Performance. In: Proceedings of AAAI Spring Sympoisum on Emotion Personality and Social Behavior (2008)
16. Thrun, S., Burgard, W., Fox, D.: Probabilistic Robotics. MIT Press (2005)
17. Torta, E., Cuijpers, R.H., Juola, J.F.: A Bayesian model for approaching a human. In: Proceedings of ECCE 2010 (2010)
18. Torta, E., Cuijpers, R.H., Juola, J.F.: A model of the user's proximity for Bayesian inference. In: Proceedings of HRI 2011 (2011)
19. Walters, M.L., Dautenhahn, K., Boekhorst, R., Koay, K.L., Syrdal, D.S., Nehaniv, C.: An empirical framework for human-robot proxemics. In: AISB 2009 (2009)

Effects of Gesture on the Perception of Psychological Anthropomorphism: A Case Study with a Humanoid Robot

Maha Salem[1], Friederike Eyssel[2], Katharina Rohlfing[2], Stefan Kopp[2], and Frank Joublin[3]

[1] Research Institute for Cognition and Robotics, Bielefeld University, Germany
msalem@cor-lab.uni-bielefeld.de
[2] Center of Excellence Cognitive Interaction Technology, Bielefeld University, Germany
{feyssel,rohlfing,skopp}@cit-ec.uni-bielefeld.de
[3] Honda Research Institute Europe, Offenbach, Germany
frank.joublin@honda-ri.de

Abstract. Previous work has shown that gestural behaviors affect anthropomorphic inferences about artificial communicators such as virtual agents. In an experiment with a humanoid robot, we investigated to what extent gesture would affect anthropomorphic inferences about the robot. Particularly, we examined the effects of the robot's hand and arm gestures on the attribution of typically human traits, likability of the robot, shared reality, and future contact intentions after interacting with the robot. For this, we manipulated the non-verbal behaviors of the humanoid robot in three experimental conditions: (1) no gesture, (2) congruent gesture, and (3) incongruent gesture. We hypothesized higher ratings on all dependent measures in the two gesture (vs. no gesture) conditions. The results confirm our predictions: when the robot used gestures during interaction, it was anthropomorphized more, participants perceived it as more likable, reported greater shared reality with it, and showed increased future contact intentions than when the robot gave instructions without using gestures. Surprisingly, this effect was particularly pronounced when the robot's gestures were partly incongruent with speech. These findings show that communicative non-verbal behaviors in robotic systems affect both anthropomorphic perceptions and the mental models humans form of a humanoid robot during interaction.

Keywords: Multimodal Interaction and Conversational Skills, Non-verbal Cues and Expressiveness, Anthropomorphism.

1 Introduction

Social robotics research is dedicated to designing, developing, and evaluating robots that can engage in social environments in a way that is both appealing and intuitive to human interaction partners. Therefore, a social robot's behavior ideally should appear natural, comprehensive, and potentially humanlike. For this, an appropriate level of communicative functionality is required which, in

B. Mutlu et al. (Eds.): ICSR 2011, LNAI 7072, pp. 31–41, 2011.

turn, strongly depends on the appearance of the robot and attributions thus made to it. Anthropomorphic design, i.e., equipping the robot with a head, two arms, and two legs, is broadly recommended to support an intuitive and meaningful interaction with humans [3,4]. It is also considered a useful means to elicit the broad spectrum of responses that humans typically direct toward each other [1]. This phenomenon is referred to as *anthropomorphism*, i.e., the attribution of human qualities to non-living objects [4]. Humanlike body features in a robot increase anthropomorphism, especially when accompanied by social-communicative movements such as gaze behavior or hand and arm gesture. But to what extent are anthropomorphic inferences determined by the robot's physical appearance and what role, on the other hand, does the robot's non-verbal behavior play with regard to judgments of anthropomorphism?

Representing a key feature of social-communicative behavior, co-verbal arm and hand gestures are primary candidates for extending the communicative capabilities of social robots. Frequently used by human speakers during interaction, gesture helps to convey information which cannot be conveyed by means of verbal communication alone, such as referential, spatial or iconic information. But gesture also affects human listeners in an interaction, as they have been shown to pay close attention to information conveyed via such non-verbal behaviors [6]. Accordingly, humanoid robots that shall be applied as interaction partners should generate co-verbal gestures for comprehensible and believable behavior. Since a large body of research (e.g., [12]) has already focused on the role of robot gaze in human-robot interaction (HRI), our investigations concentrate on hand an arm gesture as a specific subpart of non-verbal communicative behavior.

The present work aims at shedding light on how the implementation of humanlike non-verbal behaviors, specifically hand and arm gestures, affect social perceptions of the robot and HRI. For this purpose, we conducted an experiment using the Honda humanoid robot as an interaction partner. Since this robot prototype lacks visible facial features that could potentially enrich the interaction with human users (e.g., by conveying emotional states of the system), this emphasizes the necessity to rely on additional communication channels such as gestural behaviors. Therefore, we addressed this issue in the current experiment by investigating how gesture behavior would affect anthropomorphic inferences about the humanoid robot, particularly with regard to the attribution of typically human traits, likability, shared reality with the robot and judgments of acceptance, as well as future contact intentions after interacting with the robot.

2 Related Work

A large body of work (e.g., [10,2]) has evaluated complex gesture models for the animation of virtual characters. Several recent studies have investigated the human attribution of naturalness to virtual agents. In one such study [10], the conversational agent Max communicated by either utilizing a set of co-verbal gestures alongside speech, typically by self-touching or movement of the eyebrows, or by utilizing speech alone without any accompanying gestures. Participants subsequently rated Max's current emotional state and its personality, e.g., by indicating the extent to which Max appeared aggressive or lively. The results of the study showed that virtual agents are perceived in a more positive light when they produce co-verbal gestures rather than using speech as the only modality.

Despite the relevant implications of such studies, it is difficult to transfer findings from virtual agents to robot platforms. First, the presence of real physical constraints may influence the perceived level of realism. Second, given a greater degree of embodiment, interaction with a robot is potentially richer. Since humans share the same interaction space with the robot, they can walk around or even touch a real robot in an interaction study. As a result, the interaction experience is different, which is expected to affect the outcome of the results.

In the area of of human-robot interaction, much research, e.g., carried out by Mutlu et al. [12], has studied the effect of robot gaze as an important aspect of non-verbal behavior. In contrast, not much research has focused on hand and arm gestures in particular and the evaluation of their effect in HRI studies. For this reason, our work centers on speech-accompanying arm movements. However, given the strong correlation between gaze and hand gesture behavior in human communication, the interplay between these two non-verbal communication modalities needs to be further investigated in the future.

Our approach is theoretically based on social psychological research on the (de-)humanization of social groups [7]. To illustrate, [7] have proposed two distinct senses of humanness at the trait level. Specifically, they differentiate *uniquely human* and *human nature* traits. While 'uniquely human' traits imply higher cognition, civility, and refinement, traits indicating 'human nature' involve emotionality, warmth, desire, and openness. Since the human nature dimension is typically used to measure 'mechanistic dehumanization[1]', we conversely employ this measure to assess the extent to which a robot is perceived as humanlike. We further assess the degree of anthropomorphism attributed to the humanoid robot by measuring participants' perceptions of the robot's likability, shared reality with the robot, and future contact intentions.

By adapting measures of anthropomorphism from social psychological research on human nature traits [7,11], we complement existing work on the issue of measurement of anthropomorphism in social robotics (see [1] for a review). Thus, by presenting a social psychological perspective on anthropomorphism and new possible ways of measurement to the HRI community, we contribute to a deeper understanding of determinants and consequences of anthropomorphism.

In the following, we will present an experiment that tested the effects of unimodal vs. multimodal communication behavior on perceived anthropomorphism and likability, experienced shared reality, and contact intentions with regard to the robot.

[1] According to [7], characteristics representing the denial of human nature yield an image of others as being object-like or robotic. That means, when people are denied human nature, they are implicitly or explicitly objectified or likened to machines rather than to animals or humans.

3 Method

To gain a deeper understanding of how communicative robot gesture might impact and shape user experience and evaluation of human-robot interaction, we conducted a between-subjects experimental study using the Honda humanoid robot. For this, an appropriate scenario for gesture-based HRI was designed and benchmarks for the evaluation were identified. The study scenario comprised a joint task that was to be performed by a human participant in collaboration with the humanoid robot. In the given task, the robot referred to various objects by utilizing either unimodal (speech only) or multimodal (speech and either congruent or incongruent accompanying gesture) utterances, based on which the participant was expected to perceive, interpret, and perform an according action. Data documenting the participant's experience was collected after task completion using a questionnaire.

3.1 Hypothesis

We predicted that participants who received multimodal instructions from the robot (using speech and either congruent or incongruent gesture) would anthropomorphize the robot more than those who are presented with unimodal information by the robot (using only speech).

3.2 Materials

Participants interacted with the Honda humanoid robot (year 2000 model)[8]. Its upper body comprises a torso with two 5DOF arms and 1DOF hands, as well as a 2DOF head. To control the robot, we used a previously implemented speech-gesture generation model which allows for a real-time production and synchronization of multimodal robot behavior [13]. The framework combines conceptual representation and planning with motor control primitives for speech and arm movements of a physical robot body. To ensure minimal variability in the experimental procedure, the robot was partly controlled using a Wizard-Of-Oz technique during the study. The experimenter initiated the robot's interaction behavior from a fixed sequence of pre-determined utterances, each of which was triggered when the participant stood in front of the robot. Once triggered, a given utterance was generated autonomously at run-time. The ordering of the utterance sequence remained identical across conditions and experimental runs. The robot's speech was identical across conditions and was generated using the text-to-speech system *__M__odular __A__rchitecture for __R__esearch on speech s__Y__nthesis* (MARY)[14] set to a neutral voice. The entire interaction was filmed by three video cameras from different angles, while the experimenter observed and controlled the interaction from the adjacent room.

3.3 Experimental Design

The experiment was set in a simulated kitchen environment in a robot lab (see Fig. 1). The humanoid robot served as a household assistant. Participants were told that their task was to help a friend who was moving house. They were asked to unpack a cardboard box containing nine kitchen items and to put these into the cupboard that was part of the kitchen set-up. It was unknown to participants where they were supposed to put these items. However, they were informed that the robot would help them solve the task by telling them where to put the respective kitchenware. The experimental setting is illustrated in Fig. 1.

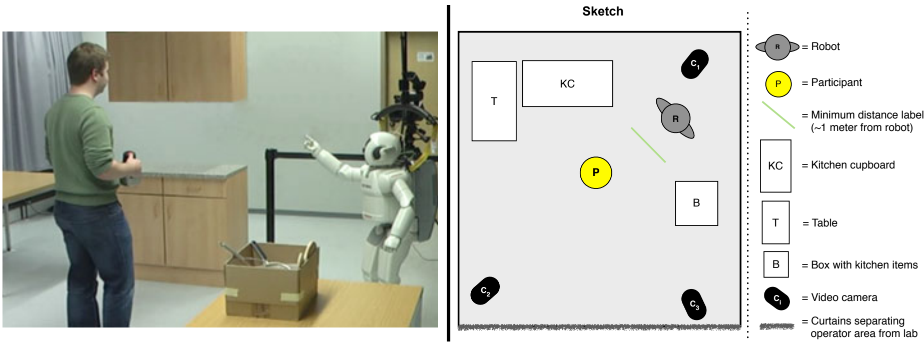


Fig. 1. The experimental setting: the robot provides the participant with information about the storage location of the object (left); sketch of the experimental lab (right)

Conditions: We manipulated the non-verbal behaviors that were displayed by the humanoid robot in three experimental conditions:

1. In the *unimodal (speech-only)* condition, the robot presented the participant with a set of nine verbal instructions to explain where each object should be placed. The robot did not move its body during the whole interaction; no gesture or gaze behaviors were performed.
2. In the *congruent multimodal (speech-gesture)* condition, the robot presented the participant with the identical set of nine verbal instructions used in condition 1. In addition, they were accompanied by a total of 21 corresponding gestures explaining where each object should be placed. Speech and gesture were semantically matching, e.g., the robot said "put it up there" and pointed up. Simple gaze behavior supporting hand and arm gestures (e.g., looking right when pointing right) was displayed during the interaction.
3. In the *incongruent multimodal (speech-gesture)* condition, the robot presented the participant with the identical set of nine verbal instructions used in condition 1. Again, in addition, they were accompanied by a total of 21 gestures, out of which ten gestures (47.6 %) semantically matched the verbal instruction, while the remaining eleven gestures (52.4 %) were semantically non-matching, e.g., the robot occasionally said "put it up there" but pointed downwards. Simple gaze behavior supporting hand and arm gestures (e.g., looking right when pointing right) was displayed during the interaction.

The *incongruent multimodal* condition was designed to decrease the reliability and task-related usefulness of the robot's gestures. In other words, participants in this group were unlikely to evaluate the use of the additional gesture modality solely based on its helpfulness in solving the given task. The choice to combine semantically matching gestures with non-matching ones in this condition was made to avoid a complete loss of the robot's credibility after a few utterances.

Verbal Utterance: In order to keep the task solvable in all three experimental conditions, spoken utterances were designed in a self-sufficient way, i.e., gestures used in the multimodal conditions were supplementary to speech. Each instruction presented by the robot typically consisted of two or three so-called *utterance chunks*. Based on the definition provided in [9], each *chunk* refers to a single idea unit represented by an intonation phrase and, optionally in a multimodal utterance, by an additional co-expressive gesture phrase. The verbal utterance chunks were based on the following syntax:

- **Two-chunk utterance:**
  `<Please take the [object]> <and place it [position+location].>`
  Example: *Please take the vase and place it on the left side of the lower cupboard.*
- **Three-chunk utterance:**
  `<Please take the [object],> <then open the [location],>`
  `<and place it [position].>`
  Example: *Please take the eggcup, then open the right drawer, and place it inside.*

Gestures: In the multimodal conditions, the robot used three different types of gesture along with speech to indicate the designated placement of each item:

- **Deictic gestures**, e.g., to indicate positions and locations
- **Iconic gestures**, e.g., to illustrate the shape or size of objects
- **Pantomimic gestures**, e.g., hand movement performed when opening cupboard doors or using a ladle

3.4 Experimental Procedure

Participants were tested individually. First, they received experimental instructions in written form. Subsequently, they entered the robot lab, where the experimenter orally provided the task instructions. They were then given the opportunity to ask any clarifying questions before the experimenter left the participant to begin the interaction with the robot. At the beginning of the experiment, the robot greeted the participant and introduced the task before commencing with the actual instruction part. The robot then presented the participant with individual utterances as described in the experimental design. Each utterance was delivered in two parts: the first part referred to the object (e.g., "*Please take the thermos flask*"); the second part comprised the item's designated position and location (e.g., "*...and place it on the right side of the upper cupboard.*"). Whenever the participant resumed a standing position in front of the robot in order to signal readiness to proceed with the next instruction, the experimenter sitting at a control terminal triggered the robot's subsequent behavior. The participant then followed the uttered instruction and, ideally, placed each item into its correct location. As explained in the briefing prior to the experimental task, participants were requested to place the object on a table adjacent to the kitchen cupboard if unsure about the item's designated location, rather than trying to guess its final position. At the end of the interaction, the robot thanked the participant for helping and bid them farewell. Participants interacted with the robot for approximately five minutes. In the unimodal (speech-only) condition all utterances including the greeting and farewell were presented verbally; in the multimodal (speech-gesture) conditions, all utterances including the greeting and farewell were accompanied by co-verbal gestures. After interacting with the robot, participants were led out of the lab to complete a post-experiment questionnaire to evaluate the robot and the interaction experience. Upon completion of the questionnaire, participants were carefully debriefed about the purpose of the experiment and received a chocolate bar as a thank-you before being dismissed.

3.5 Dependent Measures

We asked participants to report the degree to which they anthropomorphized the robot by using various dimensions. First, we measured perceived humanlikeness of the robot based on Haslam et al.'s [7] list of ten human nature traits: *curious, friendly, fun-loving, sociable, trusting, aggressive, distractible, impatient, jealous, nervous*. Second, likability was assessed using three dimensions: *polite, sympathetic, humble*. We further evaluated participants' degree of shared reality with the robot based on three items: "*How close do you feel to the robot?*", "*How pleasant was the interaction with the robot for you?*", "*How much fun did you have interacting with the robot?*". The shared reality index taps perceptions of similarity and experienced psychological closeness to the robot [5]. Moreover, it covers aspects of human-robot acceptance, since participants had to indicate how much they enjoyed the interaction with the robot. Finally, we measured participants' future contact intentions with regard to the robot using a single item: "*Would you like to live with the robot?*". All responses were given on 5-point Likert scales, with endpoints 1 = *not at all* and 5 = *very much*.

3.6 Participants

A total of 62 participants (32 female, 30 male) took part in the experiment, ranging in age from 20 to 61 years ($M = 30.90$ years, $SD = 9.82$). All participants were German native speakers and were recruited at Bielefeld University, Germany. Based on five-point Likert scale ratings, participants were identified as having negligible experience with robots ($M = 1.35$, $SD = 0.66$) and moderate skills regarding technology and computer use ($M = 3.74$, $SD = 0.97$). Participants were randomly assigned to one of three experimental conditions that manipulated the robot's non-verbal behaviors.

4 Results

First, reliability analyses (Cronbach's α) were conducted to assess the internal consistencies of the dependent measures where applicable. The indices proved sufficiently reliable, given a Cronbach's α of .78 for the index reflecting 'human nature' traits, a Cronbach's α of .73 for the 'likability' index, and a Cronbach's α of .78 for the 'shared reality' index respectively. Consequently, participants' responses to the respective items were averaged to form indices of anthropomorphism, likability, and shared reality. To test the effect of experimental conditions on the dependent measures, we conducted analyses of variance (ANOVA) and post-hoc Tukey's HSD tests with a 95% confidence interval (CI) for pairwise comparisons between condition means to test the hypothesis.

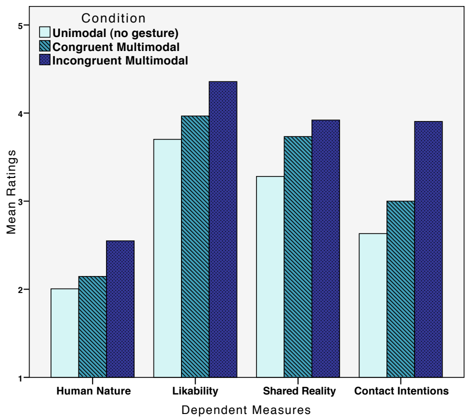
Results show a significant effect of condition on all dependent measures. Specifically, they confirm that the manipulation of the robot's gestural behavior had a significant effect on participants' attribution of human nature traits to the robot ($F(2{,}58) = 4.63$, $p = 0.01$) as well as on their assessment of the robot's likability ($F(2{,}59) = 3.65$, $p = 0.03$). Furthermore, our analyses indicate that the manipulation of the robot's non-verbal behavior had a significant effect on participants' rating of the shared reality measure ($F(2{,}59) = 4.06$, $p = 0.02$) as well as on their future contact intentions ($F(2{,}58) = 5.43$, $p = 0.007$).

Tukey post-hoc comparisons of the three groups indicate that participants in the incongruent multimodal condition ($M = 2.55$, $SD = 0.68$, CI[2.24, 2.86]) rated the attribution of human nature traits to the robot significantly higher than participants in the unimodal condition ($M = 1.98$, $SD = 0.58$, CI[1.71, 2.25]), $p = 0.013$. Moreover, participants reported significantly greater perceived likability when interacting with the robot whose verbal utterances were accompanied by partly non-matching gestures in the incongruent multimodal condition ($M = 4.36$, $SD = 0.59$, CI[4.09, 4.63]) than when it was using only speech ($M = 3.69$, $SD = 0.97$, CI[3.24, 4.15]), $p = 0.027$. Participants also experienced significantly greater shared reality with the robot when it used multimodal behaviors that were to some extent incongruent with speech ($M = 3.92$, $SD = 0.70$, CI[3.60, 4.24]) than when it relied on unimodal communication only ($M = 3.23$, $SD = 0.93$, CI[2.80, 3.67]), $p = 0.021$. Finally, participants' assessment of whether they would like to live with the robot was also significantly higher in the condition with partially incongruent speech-accompanying gesture behavior ($M = 3.90$, $SD = 1.14$, CI[3.39, 4.42]) than in the one without any gestures at all ($M = 2.63$, $SD = 1.30$, CI[2.00, 3.26]), $p = 0.007$.

Comparisons between the unimodal and the congruent multimodal condition were not statistically significant, however, they indicate a trend of higher mean ratings for all dependent measures in the congruent multimodal group.

Similarly, comparisons between the congruent multimodal and the incongruent multimodal group were not statistically significant at $p < 0.05$, however, marginally significant differences were found with regard to participants' reported future contact intentions: ratings of whether participants would like to live with the robot were higher in the incongruent multimodal condition ($M = 3.90$, $SD = 1.14$, CI[2.32, 3.59]) than in the congruent multimodal condition group

($M = 2.95$, $SD = 1.40$, CI[3.39, 4.42]), $p = 0.05$. For the remaining dependent measures, our results throughout indicate a trend of higher mean ratings in favor of the incongruent multimodal group. Fig. 2 illustrates the results.



		M (SD)
Human Nature	**Unimodal**	1.98 (.58)
	Congr. Multimod.	2.15 (.58)
	Incongr. Multimod.	2.55 (.68)
Likability	**Unimodal**	3.69 (.97)
	Congr. Multimod.	3.92 (.81)
	Incongr. Multimod.	4.36 (.59)
Shared Reality	**Unimodal**	3.23 (.93)
	Congr. Multimod.	3.75 (.76)
	Incongr. Multimod.	3.92 (.70)
Contact Intentions	**Unimodal**	2.63 (1.30)
	Congr. Multimod.	2.95 (1.40)
	Incongr. Multimod.	3.90 (1.14)

Fig. 2. Mean ratings of dependent measures as a function of experimental condition

5 Discussion and Conclusion

We conducted an experiment to investigate how hand and arm gestures affect anthropomorphic perceptions and the mental models humans form of a humanoid robot. For this, we manipulated the non-verbal behaviors of the humanoid robot in three experimental conditions: (1) no gesture, (2) congruent gesture, and (3) incongruent gesture. We particularly focused on participants' attribution of typically human traits to the robot, likability, shared reality, as well as future contact intentions. By applying a wide range of dependent variables, we examined to what extent anthropomorphic inferences on the human's side are attributed to the design, and to what extent to the behavior of the robot. Our theory-driven approach is characterized by the application of social psychological theories of (de-)humanization [7,11] to HRI. By adapting these measures of anthropomorphism from research on human nature traits, we contribute to existing work on the issue of measurement of anthropomorphism in social robotics, and thus to a deeper understanding of determinants and consequences of anthropomorphism.

The results support our hypothesis by showing that the robot's gestural behavior tends to result in a more positive subsequent evaluation of all dependent measures by the human participants. Intriguingly though, this observation was only significant for the incongruent multimodal condition, i.e., when the robot performed hand and arm gestures that did not always semantically match the information conveyed via speech, compared to the unimodal (no gesture) condition. That means, partly incongruent multimodal behavior displayed by the robot resulted in greater anthropomorphic inference as well as a more positive perception and evaluation of the robot than in the unimodal condition. These findings exceed our hypothetical expectations: not only do they suggest that a robot, even if it occasionally makes an “inappropriate” gesture, is still more favorable than a robot that does not perform any gestures at all; they also indicate that a certain level of unpredictability in a humanoid robot (as given in our incongruent gesture condition) can actually lead to a greater attribution of human traits to the robot and a more positive HRI experience. These findings are in line with previous research on anthropomorphism and social robots [4], which suggests that implementing some form of unpredictability in a robot’s behavior can create an illusion of it being “alive”. Although this observation certainly depends on the given context and task respectively, e.g., whether or not the robot’s correct and reliable behavior is vital, it could potentially lead to a paradigm shift in the design of social robots and should be further elucidated.

Future research should also investigate the generalizability of our findings regarding anthropomorphic inferences and incongruent modalities with other robotic platforms, e.g., with non-humanoid robots. Moreover, it should systematically examine the impact of gaze behavior displayed by the robot in an isolated experimental set-up without hand and arm gesture. This way we can investigate the extent to which anthropomorphic inferences, likability, shared reality and future contact intentions are determined by the robot’s arm gestures versus gaze alone. Ideally, since this was not considered in our current experiment, the robot’s behavior should also be parameterized to adapt to the human participant’s feedback in future studies. For the time being, the present results emphasize the importance of displaying gestural behaviors in social robots as significant factors that contribute to smooth and pleasant HRI. Finally, by revealing the positive impact of the incongruent gesture condition on participants’ evaluation of the robot, our findings contribute to an advancement in HRI and give new insights into human perception and understanding of gestural machine behaviors.

Acknowledgement. The work described is supported by the Honda Research Institute Europe.

References

1. Bartneck, C., Croft, E., Kulic, D.: Measuring the anthropomorphism, animacy, likeability, perceived intelligence and safety of robots. In: Proceedings of the Metrics of Human-Robot Interaction Workshop, Technical Report 471, pp. 37–41 (2008)
2. Bergmann, K., Kopp, S., Eyssel, F.: Individualized gesturing outperforms average gesturing – evaluating gesture production in virtual humans. In: Safonova, A. (ed.) IVA 2010. LNCS, vol. 6356, pp. 104–117. Springer, Heidelberg (2010)
3. Breazeal, C.: Designing sociable robots. pp. 1–263 (2002)
4. Duffy, B.R.: Anthropomorphism and the social robot. Robotics and Autonomous Systems 42(3-4), 177–190 (2003)
5. Echterhoff, G., Higgins, E.T., Levine, J.M.: Shared reality: Experiencing commonality with others’ inner states about the world. Perspectives on Psychological Science 4, 496–521 (2009)

6. Goldin-Meadow, S.: The role of gesture in communication and thinking. Trends in Cognitive Science 3, 419–429 (1999)
7. Haslam, N., Bain, P., Loughnan, S., Kashima, Y.: Attributing and denying humanness to others. European Review of Social Psychology 19, 55–85 (2008)
8. Ltd. Honda Motor Co. The Honda Humanoid Robot Asimo, year 2000 model (2000), http://world.honda.com/ASIMO/technology/spec.html
9. Kopp, S., Wachsmuth, I.: Synthesizing Multimodal Utterances for Conversational Agents. Computer Animation and Virtual Worlds 15(1), 39–52 (2004)
10. Krämer, N., Simons, N., Kopp, S.: The Effects of an Embodied Conversational Agent's Nonverbal Behavior on User's Evaluation and Behavioral Mimicry. In: Pelachaud, C., Martin, J.-C., André, E., Chollet, G., Karpouzis, K., Pelé, D. (eds.) IVA 2007. LNCS (LNAI), vol. 4722, pp. 238–251. Springer, Heidelberg (2007)
11. Loughnan, S., Haslam, N.: Animals and androids: Implicit associations between social categories and nonhumans. Psychological Science 18, 116–121 (2007)
12. Mutlu, B., Shiwa, T., Kanda, T., Ishiguro, H., Hagita, N.: Footing in human-robot conversations: how robots might shape participant roles using gaze cues. In: HRI 2009, pp. 61–68 (2009)
13. Salem, M., Kopp, S., Wachsmuth, I., Joublin, F.: Towards an integrated model of speech and gesture production for multi-modal robot behavior. In: Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication, pp. 649–654 (2010)
14. Schröder, M., Trouvain, J.: The German Text-to-Speech Synthesis System MARY: A Tool for Research. Development and Teaching 6, 365–377 (2003)

Eight Lessons Learned about Non-verbal Interactions through Robot Theater

Heather Knight

Robotics Institute, Carnegie Mellon University,
5000 Forbes Ave, Pittsburgh, PA, 15213, USA
heatherbot@cmu.edu

Abstract. Robot Theater is a fairly new arena for researching Human Robot Interaction, however, in surveying research already conducted, we have identified eight lessons from Robot Theater that inform the design of social robots today. As an interdisciplinary field, we include examples spanning robotics researchers, acting theorists, cognitive neuroscientists, behavioral psychologists and dramaturgy literature. Lessons learned include (1) the importance of intentionality in action; (2)(3)(4) the relationship between embodiment, gesture, and emotional expression; (5) the bipolar sociability categorization between machine and agent; (6) the power of interaction partners to shape robot attributions; (7) the role of audience acknowledgement and feedback; (8) the power of humor to enhance interaction. Robotics has had a long history with the field of entertainment; even the word 'robot' comes from the 1921 Czech play 'R.U.R.' – we look forward to rigorous and continued research and cross-pollination between these domains.

Keywords: Human Robot Interaction, Entertainment Robots, Non-verbal Interaction, Social Robots, Collaborations with the Arts.

1 Introduction

This paper acts as both a survey of robot theater research and a set of lessons learned about non-verbal interactions through narrative performance. For the purposes of this document, theatrical "performance" can range from literal stage with audience to pre-meditated collisions with human environments, as in guerrilla theater or street performance. The term "narrative" refers to a scripted or improvised sequence of coherent actions (and often speech) by one or more characters resulting in the construction of a timeline-based interaction arc (storyline). Also deserving of formal investigation, but beyond the scope of this paper, are puppetry, narrative video game design, and filmmaking.

Robotics has had a long history with Theater. As reported in the Encyclopedia of Computer Science [2], "Until the middle of the 20th century, mechanically programmed automata were used solely for entertainment." In Japan in the 17th century, some of the first Karakuri ningyō robots had clockwork mechanisms and a small plate for tea. When wound up, they could traverse the floor, pause for someone to pick up

B. Mutlu et al. (Eds.): ICSR 2011, LNAI 7072, pp. 42–51, 2011.

the cup, then return to their original position. Even the word "robot" as it is currently used, was coined by Karel Čapek in a play called 'R.U.R.' in 1921.

In addition to its entertainment value, investigating machine performers has research value in developing everyday robots, namely, theater is inherently social, repeatable, and there are various test subjects sitting in the audience [5][7][15]. By combining resources and creating a better mesh between these domains, it may be possible to bootstrap the development of deeper and more effective human robot interactions, particularly in the domain of non-verbal interaction. Thus, we begin this paper by discussing why non-verbal behaviors are specifically important to robot sociability, outline related knowledge in human and robot expression, then continue to the eight lessons learned about non-verbal interaction through Robot Theater. These lessons address (1) the charm of relatable gestures; (2) how affect derives from physicality; (3) movement metaphors; (4) the import of perceived rather than true robot state; (5) the gulf between machine and agent; (6) multi-agent sociability attributions; (7) the utility of audience feedback; (8) roles for machine humor. In all, theatrical contexts and/or cross-applications have enhanced human-robot interactions.

2 Motivation: Use Theater to Improve Robot Sociability

Using the theater context and body of knowledge to bootstrap the development of effective social robots is important because non-verbal expression is key to understanding sociability. In general, nonverbal response tracking capabilities could allow for more accurate social research data as such expressions are intrinsic to natural interaction. Feyereise reports, "These measures are less intrusive than questionnaires and, perhaps, also less controlled by the subjects… Moreover, impressions given by nonverbal signals influence social interactions outside laboratory conditions as in schools, courts, business situations, and medical and counseling practices. From an applied-psychology perspective, the domain for research on nonverbal communication is unlimited." [10] Furthermore, a robot's movement and engagement pattern impact our interpretation of its intention, capability, and state. With a long history of encoding and honing expression, physical theater provides pre-processed methodologies for interpreting and communicating human non-verbal behaviors that we are beginning to test on robots.

3 Background: Non-verbal Interaction

Before investigating what theatrical contexts can teach robots about non-verbal behaviors, we must establish what non-verbal behaviors are. In different systems of behavioral analysis, nonverbal communications can be grouped by channel (face, gaze, voice, bodily expression, use of space, hand position etc.), functions or psychology [10]. In this paper, we largely reference non-verbal interactions by categorical channel to enable more repeatable and easily detectable mappings within a robot behavior system. There are certain kinds of nonverbal behaviors that carry specific affective or social meaning, as in [10][16].

As we will see below, the frequency and duration of a gesture can change its associative meaning [22], e.g., how a child picks up and manipulates a fork can have clear mappings to her excitement about the food. As developmental psychologist Eliane Noirot describes, however, "meaning = movement + form + context" [10] and explicitly highlights the importance of gesture. She also indicated the central role movement plays in establishing social meaning, e.g., what the timing or presence of that action might indicate. An interesting conclusion that this suggests to me is that narrative arc and timeline are central to social interaction. This is consistent with Cory Kidd's work with Autom [13], a simple social robot designed to help people with their fitness and dieting plans, in which the robot tracked its long term interactions with a single user, moderating social behaviors as it was neglected to repair the relationship or when the user was successful to provide contextualized praise.

Various investigations have begun to use robots to explore the rich space of non-verbal behaviors in social interactions. Cynthia Breazeal's pioneering work with the Kismet robot used cues from human tone of voice and gaze, responding with simulated facial expressions and orientating toward object of interest [4]. Researchers have also begun to look at social distances, better known as proxemics, e.g., the attentional zones described in the Robo-Receptionist project [17], or in [27], where humans give wider girth to a robot that they dislike rather than like. In [25], they use the dancing robot Keepon to investigate the effect of synchronizing the robot to human rhythms on the pairs' general social interaction. In [28], researchers assess whether people can detect seemingly unintentional social cues, called 'non-verbal leakages.' There are many more.

4 Lessons Learned through Robot Theater

In this paper, we highlight eight lessons about non-verbal interactions gleaned from investigations of Robot Theater. We curate the examples to emphasize cutting-edge perspectives, whether novel or mature, on using machine performers to address modern challenges in social robotics. These examples span robotics research, acting theory, cognitive neuroscience, behavioral psychology and dramaturgy literature.

4.1 Lesson 1 - Have a Goal: Convey Intentionality

A robot using relatable gestures from entertainment-world representations can clarify current activity goals, improving its camaraderie and communication with interaction partners. For example, in recent work [33], Willow Garage partnered with a Pixar animator to "create robot behaviors that are human readable such that people can figure out what the robot is doing, reasonably predict what the robot will do next, and ultimately interact with the robot in an effective way." Their emphasis on anticipation of action and reaction to successes and failures was suggested to improve communication of intention. Though they emphasize 'readability' in the published paper, an interesting aspect of the work was one author's impatience with the large unmoving PR2 robots that might or might not be doing anything useful. When the robot used physical animations to clarify the intention and internal state of the robot, she felt more goodwill toward the system that was, nonetheless, still blocking the hallway. Further, if it failed at a task, e.g., opening a door, after several minutes of computation, but showed an animation that it 'felt bad' about not accomplishing its goal, she suggested that coworkers might commiserate, and be less annoyed by the robot's interference in their shared space.

More subtle systems of conveying robot intentionality could be created from the gestural training of French physical theater pioneer and theorist Jacque LeCoq [22]. In one case, clearly relevant to robotics, he emphasizes the importance of relative and absolute timing profiles in "look, reach, grab" behaviors to indicate motivation, emotionality and communicate social meaning. For example, if the robot is wary of an object or unsure of how to pick it up, the 'look' phase might last longest, in contrast, if multiple robots are competing for the same object, look and reach would have rapid phases in order to block out the other robots and establish claim. By using movement profiles rather than directly imitated gestures, any kinetic robot can convey intentionality, regardless of anthropomorphic form. Humans learn to communicate verbally two years after learning to communicate through gesture, thus we have excellent resources to interpret both known and novel robot gestural expressions.

4.2 Lesson 2 - There Is No Mind without Body

Human affect expressions derive from our physicality, thus robots are uniquely capable of leveraging their embodiment to communicate on human terms. As acting theorist and neuroscientist Rhonda Blair points out in [3], "A basic truth of being human is there is no consciousness without a body." As in the previous section, expression is inherently physical. She continues, "At the heart of every performance is a complex consciousness that inhabits the entire body, in which voluntary processes are inseparable from involuntary ones and in which genetic predisposition is inseparable from personal history."

Researchers have exposed unique attributes of embodiment and physical presence in enhancing artificial agent interaction with a human, as contrasted to virtual agents (e.g. [18][21]). In [18], they use drumming games with children to contrast a physical robot to a virtual robot that mirrors the first case. Given participant explicit feedback and implicit measurements, they concluded that the "presence of a physical, embodied robot enabled more interaction, better drumming and turn-taking, as well as enjoyment of the interaction, especially when the robot used gestures." James-Lange Theory [19] is a somatic theory of emotion, in which, "the perception of bodily changes as they occur *is* the emotion."

The idea that behavior precedes feeling is important. Regardless of complete literal or physiological truth, this framing of body on higher or equal footing with the mind will allow robotics to more fully leverage its physicality to improve expression. Theorist's say "there's an affect associated with every functioning of the body, from moving your foot to taking a step to moving your lips to make words. Affect is simply a body movement looked at from the point of view of its potential — its capacity to come to be, or better, to come to do." [34] In other words, a robot not fully leveraging its physicality not only loses a mode of communication but is also less expressive.

4.3 Lesson 3 - Mirror Neurons: Physicality and Motion

We often interpret robot behaviors, especially non-verbal expressions, by re-mapping them on to ourselves, thus robots can provide people with effective stimuli given a deeper knowledge of movement metaphors. Though robots may or may not share human physiology, their actions might be simulated, as best as our brain can, through our own physical capacities. Neurophysiologist Giacomo Rizzolatti [3] describes the function of mirror neurons; "Mirror neurons allow us to grasp the minds of others, not through conceptual reasoning, but through direction simulation, by feeling, not by thinking." This has ramifications on the theater community's long-controversial analysis of the power-relationship between performer and audience. The impact of this new understanding is that watching something might be, neurologically, the same thing as doing something; the parallel neurons will fire.

Theater provides possible encodings for movement metaphors, as in the extensive work of Michael Chekhov and his collection of emotional gestures [31], which posits the body as the source of major metaphors of thought, meaning and value. Japanese Noh Theater has a related system of symbolic meanings as codified gestures, so consistent that the line between metaphor and literal significance virtually disappears [20]. "A crucial implication of [mirror neurons] is that the metaphors we live by are not just abstract or poetic, but are of our bodies in the most immediate way" [3]. Perhaps even learned gestures are, at their source, based in physiological experience.

4.4 Lesson 4 - Outward Emotional Communication Trumps Inward Experience

Our perception of a social robot's expression has more influence on interaction than its true internal state; complexity is not a prerequisite. As actress Elsie de Brauw helps explain: "Observation of what the spectator sees and what I experience as an actress, is completely different. Moreover, who sees those tears? Only the people in the first four rows." [24] As researchers, we might think first to the robot's internal model of an interaction, but dramaturgy reminds us, it is the viewers that matter most.

Not only should intentionality be clear, but even stronger than that, outward intentionality outweighs an actor's internal feelings, and the two may be out of sync. Many a performer has left the stage with a pit in his stomach, bemoaning his lack of authentic emotion, only to receive highest praise. Now, this may be a good time to highlight the difference between robots and people. As currently designed, most robots are intended to enhance, enable or empower a human or set of humans. Thus the inner experience of that robot is trumped by the success of that larger interaction loop. Similarly, an actor on stage is generally tasked with the goal of entertaining or evoking response in the audience to whom he is performing. Thus, the metaphor may provide parallel techniques that can be used across domains.

One of the most successful social robots ever created, in terms of meeting its goal, inspiring third party human-to-human interaction, encouraging goodwill toward the robot and resulting in a sense of accomplishment for the humans involved, was the Tweenbot [14]. Designer Kacie Kinzer deployed the simple robot in a in New York

City park. Its small cardboard structure had a marker-drawn smiley-face, motors that went constantly forward, and a flag declaring, “Help me. I’m trying to reach the South-West corner. Aim me in the right direction.” Forty-two minutes after release in the park with the creator out of associative view, twenty-nine diverse un-instructed strangers had read the sign and helped the robot. Just when it reached the edge of the park, one last stranger scooped it up and turned it around, saying, “You can’t go that way. It’s toward the road.” Sometimes simplicity and clean physical design can be the clearest way to streamline our communication of robot intention.

4.5 Lesson 5 - Social Role: The Gulf between Props and Character

The categorization of ‘social robot’ denotes a machine that behaves and is treated as agent. In analyzing what makes a robot on stage a categorical character (agent) versus prop (object), a sense of authentic interactions seems to distinguish these labels. We see this in the professional theater productions with robots from this past year, e.g., Todd Machover’s robot opera called ‘Death and the Powers,’ the remote-controlled robots in ‘Heddatron,’ or the casting (in a very wide definition of robot) of an enormous puppeteered equestrian structure in ‘Warhorse.’ With the exception of the final, directly human manipulated case, these robots fell into the domain of props because of the lack of believable interaction arcs with the human actors on stage.

One success story is the realistic humanoid Geminoid-F’s seated one-scene performance [11] with a single human actor. Because of the central role she played, dying of a terminal disease, and human-like features, she fell on the agent side of the gulf, but additional local intelligence and integration of audience feedback between performances could further improve our sense of her agency. In a parallel case off stage, [32] analyzed the relationship of groups of humans with a robotic photographer that was largely ignored. The contrasting settings included a conference hall and a wedding. Researchers noted a vast difference in behavioral reactions to the system depending on whether it was perceived as social (people wanted to wave at it, talk to it, especially when it accidentally displayed double-take motion), versus as an object.

Interactivity in interfaces is an old domain made new through modern technology, as explored in the MIT play I/It [30]. Modern entertainment has been passing through an unusual phase of ‘object.’ As reported by Douglas Adams in 1999, "During this century we have for the first time been dominated by non-interactive forms of entertainment: cinema, radio, recorded music and television. Before they came along all entertainment was interactive: theatre, music, sport; the performers and audience were there together, and even a respectfully silent audience exerted a powerful shaping presence on the unfolding of whatever drama they were there for. We didn't need a special word for interactivity in the same way that we don't (yet) need a special word for people with only one head"[1]. It is time to bring the interaction and agency back through artificial social intelligence. Other notable projects featuring attributions of intelligence and agency are the ability for a robot to engage in improvised generative performances [6], pose and/or answer questions [9], or achieve appropriate timing and attention orientation when acting with human actors [12].

4.6 Lesson 6 - Good Actors Outweigh Bad Actors: Attribution

Multi-robot or human-robot teams can be greater than the sum of their parts in perceived interaction intelligence. The stage provides a ripe context to explore these effects. The University of Texas rendition of 'Midsummer Night's Dream' cast small flying robots as fairies [29], finding that intentioned social actors influence third parties interactions: "Consistent with stage theory, where the visible reaction of the actor to an action by another actor creates the impression of affect, the human actors can create affect even if the robot's actions are independent."

First introductions by social actors were particularly important: "If a micro-heli crashed into the stage first and the audience saw a fairy treating the robot as a baby, the audience invariably duplicated the action. The audience member might be surprised, but not visibly annoyed, and would gently pick up the robot and hold it in their palm to allow a relaunch… However, if a micro-heli crashed into the audience first, the audience member was generally disgruntled. Observed reactions by the audience were kicking the robot back onto the stage, throwing the robot like a baseball apparently intending to relaunch it, or passing it to the end of the aisle. It was significant that the audience did not look to the operators for instruction as to what to do with the robot; the audience member seemed to look for cues on how to behave from the actors or the robot." [29]

The play also provided insight on the potentials of un-even human robot teams, "The actors compensated for the robot's lack of control and unpredictably location, creating an impression of cooperation." [29] One might imagine multi-robot teams capable of leveraging similar failure modes to maintain successful interactions with a third party.

4.7 Lesson 7 - Acknowledgement/Learning: Looping in Audience Feedback

Human audiences are already highly cognizant of human social behaviors and can provide real time feedback to robot comportment on stage, thus audience tracking in theater settings is an important new domain for experimentation. As Brook proclaims, "The audience assists the actor, and at the same time for the audience itself assistance comes back from the stage" [24]. With that motivation, I recently began a robot standup comic project in which a robot loops in audience feedback to change the content and emphasis of its performance [15]. In its first iteration, we loaded a database of pre-scripted jokes onto a Nao robot, scoring each entry along five initial attribute scales {movement level, appropriateness, likelihood to have been heard before, length, and interactivity level}.

The desired goal of the audience tracking was to maximize the audience's overall enjoyment level. In practice, the robot uses an estimate of the audience's enjoyment-level (using laugher/applause audio and red/green feedback card visuals) in reaction to the previous joke to update the robot's hypothesis of what attributes the audience likes and dislikes. We use that estimate to predict the audience's attribute enjoyment preferences, as summed up by the weight vector w(t), and increase or decrease each attributes' value by multiplying the valence of the response, y, with the characteristics of the previous joke J(t) and a learning-rate constant α. Thus audience model is updated to the next timestep, w(t+1), using the equation, $w(t+1) = w(t) + \alpha y J(t)$. In mathematical terms, this technique is called online convex programming.

Even beyond the real-time feedback, the audience continues to provide instrumental verbal and email feedback about the performance, including the attention of professional comedians to help generate new and future scripts. The sneak peak I can offer the readers here is: Never 'eat' the audience's laughter or applause by beginning the next joke too early; Acknowledge the valence of the audience response to achieve higher knowledge attribution and make them feel more included, whether verbally or through gaze and body pose; consider overarching arc; and develop a good rhythm for each individual joke. Many of these ideas generalize to social robotics.

4.8 Lesson 8 - Humor Will Make People Like Your Robot Better

Humor can enhance human-robot interaction by helping creating common ground, trust or forgiveness, but its subtlety makes collaboration with theater communities uniquely beneficial. As comedian Drucker spells out for us in this following snippet [8], robot performances can go terribly wrong (note: this is a fictional robot standup performance): "Hello, world! What level is everyone's excitement currently at? I'm sorry. I cannot hear you. Would you please repeat your excitement, preferably at a louder volume? Thank you. I am also excited. Have you ever noticed the difference between white robots and black robots? White robots are all 1001001, but black robots are all 0110110. Do you agree? You have said you do not agree."

In [26], users assigned to computers that used humor during a procedural task rated the agents as more likable, reported greater cooperation between themselves and the machine, and declared more feelings of similarity and relatablity with the system. Combining [23] with the Willow Garage experiment with Pixar [33], I suggest that if a robot not only acknowledges its failing, but also make a self-deprecating joke, people may find their interactions with a faulty robot enjoyable. Of course, humor is one of the most elusive and human of the social traits. Within that, timing is one of the most challenging considerations so perhaps we can also experiment with shared-autonomy performances as in [12].

Professional comedians and joke writers have been polishing their art for thousands of years, much like theater professionals are the experts of artificial emotion and personality. So, from personal experience, I recommend collaborating with the masters. Our social behaviors as humans developed a very long time ago, so if humor can help a robot seem like one of our 'tribe' that could be a huge leap toward overcoming the paradigm where robots are only seen as tools (or props).

5 Conclusion: The Role of Robot Theater

This paper outlines ways in which physical theater applied to robotics has already provided a deeper understanding of how intentional or coincidental robot actions might impact human perception. The new nonverbal behavior toolsets, gleaned from our survey of Robot Theater explorations, include movement profiles [19][22][34], symbolic [31] or mirrored gestural expression [3], and the use of the stage and audience as a context for testing and improving robot sociability [5][7][15], social attributions [24][29] or assessment of agency [6][11][12]. While we acknowledge that theatrical contexts are distinct from natural sociability, robotic interaction schemas generally place humans at the center of overall task goals, thus there are many overlapping lessons we can glean from the construct of an actor and audience.

In summary, we have established that (1) robots using relatable gestures can clarify current activity goals and improve camaraderie; (2) human affect expressions derive from our physicality, thus robots are uniquely capable of leveraging their embodiment to communicate on human terms; (3) theater provides encodings for movement metaphors such that robot actions might mirror onto ourselves; (4) human perception is a better benchmark for a robot's social design than internal AI; (5) a machine must convey social intelligence to make the leap from object to agent; (6) theater settings provide a unique testing ground for developing multi-agent interactions with a third party; (7) audiences provide useful visceral and conscious feedback data to social robots in development; (8) machine humor, though difficult to design, is highly impactful to interaction and a fertile domain for interdisciplinary collaboration.

We hope the social robotics community will find useful information within this sampling, and leverage our findings to motivate additional investigations. Future work should continue to evaluate cross-applications of social knowledge from dramaturgical theory to robot behavior systems, and envision contexts for Robot Theater that frame the audience as a user study full of participants.

Acknowledgments. Special thanks to CMU Acting Prof. Matthew Gray and Dramaturgy Prof. Michael Chemers for helping provide many theater and psychology references below. Thanks also to the NSF Graduate Research Fellowship Program.

References

1. Adams, D.: How to Stop Worrying and Learn to Love the Internet. Sunday Times: News Review (August 29, 1999)
2. Bhatti, R.: Robotics. In: Encyclopedia of Computer Science, John Wiley and Sons Ltd., Chichester (2003)
3. Blair, R.: Actor, Image, and Action: Acting and Cognitive Neuroscience. Routledge, New York (2008)
4. Breazeal, C.: Designing Sociable Robots. MIT Press, Cambridge (2004)
5. Breazeal, C., et al.: Interactive robot theatre. In: IEEE/RSJ IROS (2003)
6. Bruce, A., et al.: Robot Improv: Using drama to create believable agents. AAAI (2000)
7. Demers, L.: Machine Performers: Neither Agentic nor Automatic. In: ACM/IEEE HRI Workshop on Collaborations with Arts (2010)
8. Drucker, M.: A Robot Performs Standup Comedy to a Lackluster Response. McSweeney's (November 6, 2007), `http://www.mcsweeneys.net/2007/11/6drucker.html`
9. Fong, T., et al.: Robot asker of questions. Robotics and Autonomous Systems, 235–243 (2003)
10. Feyereise, P., de Lannoy, J.: Gestures and Speech: Psychological Investigations. Cambridge University Press, Cambridge (1991)

11. Guizzo, E.: Geminoid F Gets Job as Robot Actress, IEEE Automoton Blog (2011), http://spectrum.ieee.org/automaton/robotics/humanoids/geminoid-f-takes-the-stage
12. Hoffman, G., Kubat, R.: A Hybrid Control System for Puppeteering a Live Robotic Stage Actor. RoMAN (2008)
13. Kidd, C.: Designing for Long-Term Human-Robot Interaction and Application to Weight Loss. Ph.D. Dissertation, Massachusetts Institute of Technology (2008)
14. Kinzer, K.: Robot Love. Poptech (2009), http://poptech.org/popcasts/kacie_kinzer_robot_love, additional info at http://www.tweenbots.com/
15. Knight, H.: A Savvy Robot Standup Comic: Online Learning through Audience Tracking. Workshop paper, ACM TEI (2010)
16. Knight, H., et al.: Realtime Touch Gesture Recognition for a Sensate Robot. In: IEEE/RSJ IROS (2009)
17. Kirby, R., et al.: Designing Robots for Long-Term Social Interaction. In: IEEE/RSJ IROS (2005)
18. Kose-Bagci, H., et al.: Effects of Embodiment and Gestures on Social Interaction in Drumming Games with a Humanoid Robot. Advanced Robotics 23 (November 2009)
19. James, W.: What is Emotion? First Published in: Mind 9, 188–205 (1884)
20. Yamazaki, M.: On the Art of the No Drama: The Major Treatises of Zeami. Princeton University Press (1984)
21. Lee, K.M., Jung, Y., Kim, J., Kim, S.R.: Are physically embodied social agents better than disembodied social agents?: The effects of physical embodiment, tactile interaction, and people's loneliness in human-robot interaction. Int'l Journal of Human-Computer Studies (2006)
22. LeCoq, J.: The Moving Body: Teaching Creative Theater. Routledge, New York (2002)
23. Lee, M.K., Kiesler, S., Forlizzi, J., Srinivasa, S., Rybski, P.: Gracefully mitigating breakdowns in robotic services. In: ACM/IEEE HRI (2010)
24. Konijn, E.A.: Acting Emotions. Amsterdam University Press, Amsterdam (2000)
25. Michalowski, M., Sabanovic, S., Kozima, H.: A dancing robot for rhythmic social interaction. In: ACM/IEEE HRI (2007)
26. Morkes, J., Kernal, H., Nass, C.: Effects of Humor in Task-Oriented Human-Computer Interaction and Computer-Mediated Communication. Journal of Human-Computer Interaction (December 1999)
27. Mumm, J., Mutlu, B.: Human-Robot Proxemics: Physical and Psychological Distancing in Human Robot Interaction. In: ACM/IEEE HRI (2011)
28. Mutlu, B.: Nonverbal Leakage in Human Robot Interaction. In: ACM/IEEE HRI (2009)
29. Murphy, R., et al.: A Midsummer Night's Dream (with Flying Robots). In: ACM/IEEE HRI Workshop on Collaborations with the Arts (2010)
30. Pinhanez, C., Bobick, F.: It/I: a theater play featuring an autonomous computer character. MIT Press Journals: Presence, Teleoperators and Virtual Environments (October 2002)
31. Rumohr, F.: Movement for Actors. In: Potter, N. (ed.) Michael Chekhov, Psychological Gesture, and the Thinking Heart. Allworth Press, New York (2002)
32. Smart, W., Grimm, C., Doxin, M., Byers, Z.: (Not) Interacting with a Robot Photographer. AAAI (2003)
33. Takayama, L., Dooley, D., Ju, W.: Expressing thought: Improving robot readability with animation principles. In: ACM/IEEE HRI (2011)
34. Zournazi, M.: Navigating Movements: An Interview with Brian Massumi. 21C Magazine (2003)

Proxemic Feature Recognition for Interactive Robots: Automating Metrics from the Social Sciences

Ross Mead, Amin Atrash, and Maja J. Matarić

Interaction Lab, Computer Science Department, University of Southern California
{rossmead,atrash,mataric}@usc.edu

Abstract. In this work, we discuss a set of metrics for analyzing human spatial behavior (proxemics) motivated by work in the social sciences. Specifically, we investigate individual, attentional, interpersonal, and physiological factors that contribute to social spacing. We demonstrate the feasibility of autonomous real-time annotation of these spatial features during multi-person social encounters. We utilize sensor suites that are non-invasive to participants, are readily deployable in a variety of environments (ranging from an instrumented workspace to a mobile robot platform), and do not interfere with the social interaction itself. Finally, we provide a discussion of the impact of these metrics and their utility in autonomous socially interactive systems.

Keywords: Proxemics, spatial interaction, spatial dynamics, social spacing, social robot, human-robot interaction, PrimeSensor, Microsoft Kinect.

1 Introduction and Background

Proxemics is the study of the dynamic process by which people position themselves in face-to-face social encounters [1]. This process is governed by sociocultural norms that, in effect, determine the overall sensory experience of each interacting participant [2]. People use proxemic signals, such as distance, stance, hip and shoulder orientation, head pose, and eye gaze, to communicate an interest in initiating, accepting, maintaining, terminating, or avoiding social interaction [3-5]. People can also manipulate space in an interaction, perhaps to direct attention to an external stimulus (usually accompanied by a hand gesture) or to guide a social partner to another location [6]. These cues are often subtle and noisy, and, subsequently, are subject to coarse analysis.

There exists a considerable body of work in the social sciences that seeks to analyze and explain certain proxemic phenomena. The anthropologist Edward T. Hall [1] coined the term "proxemics", and proposed that *physiological* influences shaped by culture define zones of proxemic distances [2, 7]. Mehrabian [5], Argyle and Dean [8], and Burgoon *et al.* [9] analyzed spatial behaviors as a function of the *interpersonal* relationship between social partners. Schöne [10] was inspired by the spatial behaviors of biological organisms in response to stimuli, and investigated human spatial dynamics from physiological and *ethological* perspectives; similarly, Hayduk

B. Mutlu et al. (Eds.): ICSR 2011, LNAI 7072, pp. 52–61, 2011.

and Mainprize [11] analyzed the personal space requirements of people who are blind. Kennedy *et al.* [12] studied the amygdala and how *emotional* (specifically, fight-or-flight) responses regulate space. Kendon [13] analyzed the *organizational* patterns of social encounters, categorizing them into *F-formations*: "when two or more people sustain a spatial and orientation relationship in which the space between them is one to which they have equal, direct, and exclusive access." Proxemic behavior is also impacted by factors of the *individual*—such as sex [14], age [15], ethnicity [16], and personality [17]—as well as *environmental* features—such as lighting [18], setting [19], location in setting and crowding [20], size [21], and permanence [7].

Following the emergence of embodied conversational agents [22], the study of proxemics was approached from a *computational* perspective. Rule-based social formation controllers have been applied to human-robot interaction (HRI) [23-25]. Interpersonal dynamic theories, such as equilibrium theory [8], have been evaluated in HRI [26, 27] and immersive virtual social environments [28]. Contemporary machine learning techniques have been applied to socially-appropriate person-aware navigation in dynamic crowded environments [29] and recognition of positive and negative attitudes of children with autism to an interactive robot [30].

A lack of high-resolution metrics limited previous efforts to coarse analyses in both space and time [31, 32]. Recent developments in markerless motion capture, such as the Microsoft Kinect[1] and the PrimeSensor[2], have addressed the problem of real-time human pose estimation, providing the means and justification to revisit and more accurately model the subtle dynamics of human spatial interaction. In this work, we present a system that takes advantage of these advancements and draws on inspiration from existing metrics in the social sciences to automate the analysis process of proxemic behavior. This automation is necessary for the development of socially-situated artificial agents, both virtual and physically embodied.

2 Metrics for Proxemic Behavior Analysis

In this paper, we consider metrics that are commonly used by the social sciences to analyze proxemic behavior. Specifically, we are interested in the validated methods employed by Schegloff [33], McNeill [6, 34], Mehrabian [5], and Hall [35].

2.1 Schegloff's Metrics

Schegloff [33] emphasized the importance of distinguishing between relative poses of the lower and upper parts of the body (Fig. 1), suggesting that changes in the lower parts (from the waist down) signal "dominant involvement", while changes in the upper parts (from the waist up) signal "subordinate involvement". He noted that, when a pose would deviate from its "home position" (i.e., 0°) with respect to an adjacent pose, the deviation would not last long and a compensatory orientation behavior

[1] `http://www.xbox.com/en-US/kinect`

[2] `http://www.primesense.com/`

would occur, either from the subordinate or the dominant body part. More often, the subordinate body part (e.g., head) is responsible for the deviation and, thus, provides the compensatory behavior; however, if the dominant body part (e.g., shoulder) is responsible for the deviation or provides the compensatory behavior, a shift in attention (or "involvement") is likely to have occurred. Schegloff referred to this phenomenon as "body torque".

- **Stance Pose:** most dominant involvement cue; position midway between the left and right ankle positions, and orientation orthogonal to the line segment connecting the left and right ankle positions
- **Hip Pose:** subordinate to stance pose; position midway between the left and right hip positions, and orientation orthogonal to the line segment connecting the left and right hip positions
- **Torso Pose:** subordinate to hip pose; position of torso, and average of hip pose orientation and shoulder pose orientation (weighted based on relative torso position between to hip pose and shoulder pose)
- **Shoulder Pose:** subordinate to torso pose; position midway between the left and right shoulder positions, and orientation orthogonal to the line segment connecting the left and right shoulder positions
- **Hip Torque:** angle between the hip and stance poses
- **Torso Torque:** angle between the torso and hip poses
- **Shoulder Torque:** angle between the shoulder and torso poses

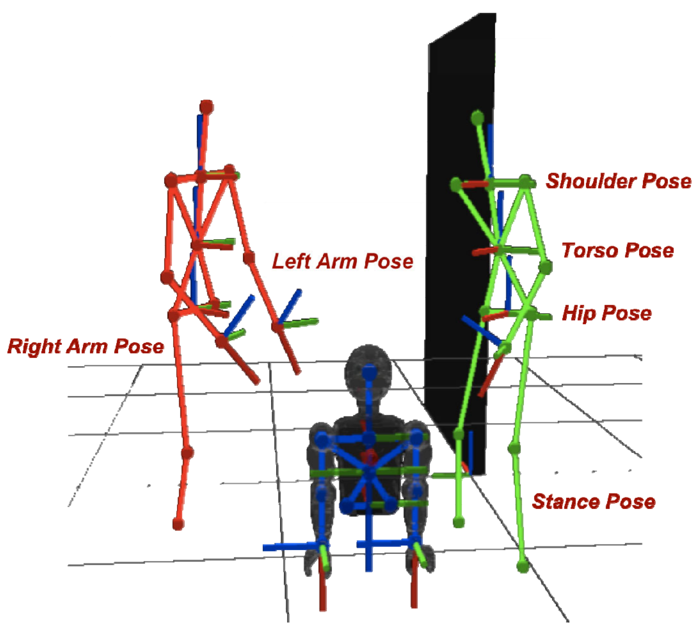


Fig. 1. Pose data for the two human users and an upper torso humanoid robot; the absence of some features, such as arms or legs, signified a position estimate with low confidence

2.2 McNeill's Metrics

McNeil [6, 34] posited that *attentional* (or *instrumental*) gestures could be used to manipulate space in social interactions by incorporating an object of mutual interest. As a first step to recognizing these features, we considered the relationship between the elbow and the hand of each individual in the interaction (Fig. 1).

- **Left Arm Pose:** position of left elbow, and orientation from left elbow to left hand
- **Right Arm Pose:** position of right elbow, and orientation from right elbow to right hand
- **Left Arm Deictic Referent:** target (social agent or stimulus) of a ray projected from the left elbow through the left hand
- **Right Arm Deictic Referent:** target (social agent or stimulus) of a ray projected from the right elbow through the right hand

2.3 Mehrabian's Metrics

Mehrabian [5] provides distance- and orientation-based interpersonal metrics for proxemic behavior analysis (Fig. 2). These spatial features are the most commonly used in the study of proxemics in both human-human and human-robot interactions.

- **Total Distance:** magnitude of a *Euclidean interpersonal distance vector* between the closest two points (one from each individual) in a social dyad
- **Straight-Ahead Distance:** magnitude of the *x*-component of the distance vector
- **Lateral Distance:** magnitude of the *y*-component of the distance vector
- **Relative Body Orientation:** magnitude of the angle between a social dyad

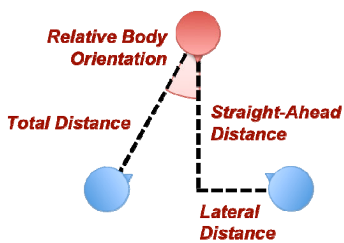


Fig. 2. In this triadic (three individual) interaction scenario, proxemic behavior is analyzed using simple spatial metrics between each social dyad (pair of individuals)

2.4 Hall's Metrics[3]

Hall's physiologically-inspired proxemic metrics are proposed as an alternative to strict spatial analysis, providing a sort of functional sensory explanation to the human use of space in social interaction [35] (Fig. 3). Hall seeks not only to answer questions of *where* and *when* a person will be, but, also, the question of *why*.

[3] For completeness, Hall's sex code (male or female) and posture code (prone, sitting, squatting, or standing) were also recorded by hand; automatic recognition of these proxemic factors was not attempted.

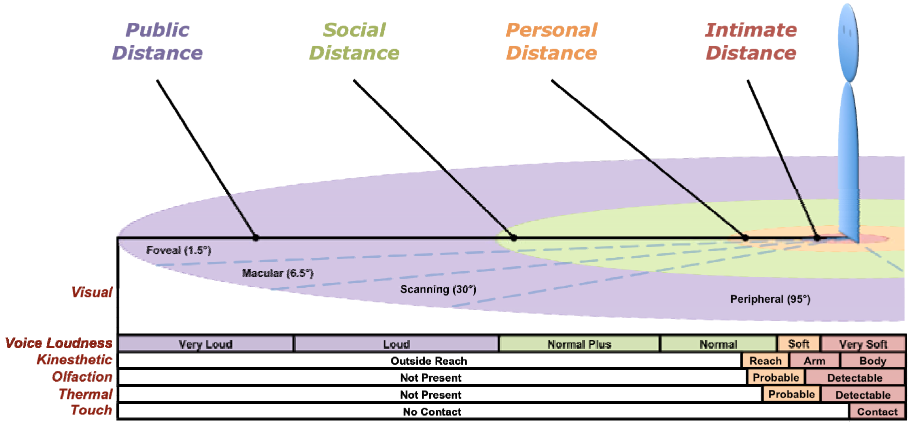


Fig. 3. Public, social, personal, and intimate distances, and the anticipated sensory sensations that an individual would experience while in each of these proximal zones

- **Distance Code[4]:** based on total distance; close-phase intimate distance (0”-6”), far-phase intimate distance (6”-18”), close-phase personal distance (18”-30”), far-phase personal distance (30”-48”), close-phase social distance (48”-84”), far-phase social distance (84”-144”), close-phase public distance (144”-300”), or far-phase public distance (more than 300”)
- **Visual Code:** based on relative body orientation[5]; foveal (sharp; 1.5° off-center), macular (clear; 6.5° off-center), scanning (30° off-center), peripheral (95° off-center), or no visual contact
- **Voice Loudness Code:** based on total distance; silent (0”-6”), very soft (6”-12”), soft (12”-30”), normal (30”-78”), normal plus (78”-144”), loud (144”-228”), or very loud (more than 228”)
- **Kinesthetic Code:** based on the distances between the hip, torso, shoulder, and arm poses; within body contact distance, just outside body contact distance, within easy touching distance with only forearm extended, just outside forearm distance (“elbow room”), within touching or grasping distance with the arms fully extended, just outside this distance, within reaching distance, or outside reaching distance
- **Olfaction Code:** based on total distance; differentiated body odor detectable (0”-6”), undifferentiated body odor detectable (6”-12”), breath detectable (12”-18”), olfaction probably present (18”-36”), or olfaction not present
- **Thermal Code:** based on total distance; conducted heat detected (0”-6”), radiant heat detected (6”-12”), heat probably detected (12”-21”), or heat not detected

---

[4] These proxemic distances pertain to Western American culture—they are not cross-cultural.

[5] In this implementation, shoulder pose was used to estimate the visual code, as the size of each person’s face in the recorded image frames was too small to recognize eye gaze or head orientation.

- **Touch Code:** based on total distance; contact[6] or no contact
- **Sociofugal-Sociopetal (SFP) Axis Code:** based on relative body orientation (in 22.5° intervals), with face-to-face (axis-0) representing maximum sociopetality and back-to-face (axis-8) representing maximum sociofugality [36-38]; axis-0 (0°-11.25°), axis-1 (11.25°-33.75°), axis-2 (33.75°-56.25°), axis-3 (56.25°-78.75°), axis-4 (78.75°-101.25°), axis-5 (101.25°-123.75°), axis-6 (123.75°-146.25°), axis-7 (146.25°-168.75°), or axis-8 (168.75°-180°)

3 Pilot Study

We conducted a pilot study to observe and analyze human spatial dynamics in natural interactions. The objective of this study was to demonstrate the feasibility of real-time annotation of proxemic behaviors in multi-person social encounters. In particular, we are interested in working with sensor suites that (1) are non-invasive to participants, (2) are readily deployable in a variety of environments (ranging from an instrumented workspace to a mobile robot), and (3) do not interfere with the interaction itself.

The study was set up and conducted in a twenty-by-twenty room in the Interaction Lab at the University of Southern California. A “presenter” and a participant engaged in a 5-6 minute open-ended (i.e., non-scripted) interaction loosely focused on a common object of interest—a static, non-interactive humanoid robot. (In this work, the second author played the role of the presenter, though was unaware of the conditions of the experiment.) The interactees were monitored by the PrimeSensor markerless motion capture system, an overhead color camera, and an omnidirectional microphone. A complete description of the experimental setup can be found in [39].

4 Qualitative Analysis and Discussion

A total of 18 participants were involved in the study. Joint positions recorded by the PrimeSensor were processed to recognize features based on the metrics of Schegloff [33], McNeill [6, 34], Mehrabian [5], and Hall [35], each with corresponding confidence values. We calculated these proxemic metrics and their respective changes for the participant, the presenter, and the robot in real-time. A preliminary validation of the feature recognition is available in [40]; a comprehensive analysis of system performance is currently underway.

For Schegloff’s metrics [33], pose information was accurately extracted from the PrimeSensor data [40]. Each joint position provided a confidence value, which was used to estimate the quality of the recognition of each individual pose. In addition, these poses simplified distance and orientation estimates for Mehrabian’s metrics [5].

To identify McNeill’s metrics [6], a ray was extended from the arm of the individual and used to detect collisions against any objects in the environment. This naïve

[6] More formally, Hall’s touch code distinguishes between caressing and holding, feeling or caressing, extended or prolonged holding, holding, spot touching (hand peck), and accidental touching (brushing); however, automatic recognition of such forms of touching go beyond the scope of this work.

approach proved limited in detecting pointing behaviors in more complex social interactions—overall, the processed deictic referents were too noisy to be useful [40]. In future work, a model for recognizing pointing gestures will be trained to discriminate deictic behavior from other social and instrumental activities [34].

To our knowledge, this is the first time that Hall's physiologically-inspired proxemic metrics [35] have been automatically extracted. By utilizing the distances between joint estimates, kinesthetic factors were accurately modeled for each individual [40]. Basic 3D collision detection techniques were used to identify contact between the participants (e.g., during a handshake). With the utilized sensor technologies, it would be feasible to consider a continuous measure of orientation, as opposed to the discretized SFP axis code. However, the sensors used in the current work were unable to provide eye gaze or head orientation; thus, metrics such as visual and olfactory receptor fields were estimated based on the recognized shoulder poses of the individuals. Future work will utilize additional sensors to estimate these metrics; for example, the Watson 3D head orientation tracker [41] could be used to estimate the field-of-view and olfactory fields, temperature and wind sensors to estimate thermal and olfactory factors, and more sensitive microphones to measure voice loudness. The addition of these sensors would provide the means to analyze spatial interaction dynamics in ways never before possible; improved recognition techniques via such sensors would improve their usability for autonomous systems. Finally, our anecdotal observations suggest that sensory (i.e., auditory, visual, thermal, olfactory) interference from the environment could extend the defined metrics and provide significant insight into the study of proxemic behavior [39]. In such cases, Hall's metrics [35] might serve as better representations of the interaction than simple spatial metrics typically utilized in social encounters. Such an approach has yet to be investigated by any computational model, and will be a distinguishing factor in our continued work.

5 Conclusions and Future Work

In this work, we presented a set of metrics motivated by work in the social sciences for analyzing human spatial behavior. We then described a pilot study to demonstrate the feasibility of the real-time annotation of these spatial metrics in multi-person social encounters [39]. We utilized a markerless motion capture system (the PrimeSensor), an overhead camera, and an omnidirectional microphone to recognize these metrics in real-time; this sensor suite is non-invasive to participants, is readily deployable in a variety of environments, and does not interfere with the interaction itself.

We are currently in the process of validating our automated system by comparing the performance against multiple human annotators; a preliminary validation is available in [40]. Our data set includes approximately 2 hours of data with 13 features per individual and 16 features per dyad. Although it is difficult to estimate the actual time for labeling by annotators, protocol within the USC Interaction Lab suggests allocating one hour per one minute of data for each feature; this further highlights the utility an automated annotation system. As future work focuses on collection of larger datasets, it becomes prohibitive to use human annotators to label the data. Additionally, as autonomous agents use these metrics to conduct social interaction, it is important to understand the accuracy of the system for decision-making. An automated recognition system must provide sufficient accuracy to be useful.

We now present future directions and applications of this work, particularly with applications to autonomous systems.

5.1 Heuristic-Based vs. Learned Approaches

Currently, the models used by our automated system utilize heuristics based on metrics provided by the literature [5, 6, 33-35]. These parameters are often determined from empirical evidence as observed by the researchers. Consequently, this approach results in a discretization of the parameter space. However, advances in machine learning techniques allow for parameters to be determined automatically from data. This provides several advantages. First, it allows for continuous representation of the data. Current approaches discretize the parameter space into large regions. Automated systems can focus on learning continuous values of the data, resulting in a more fine-grained representation. Second, automatic learning allows for integration of large datasets; by taking advantage of automated labeling, models can be constructed from extensive datasets, which might be more accurate than hand-labeled models built from small datasets. Finally, learning of the models allows for impromptu and dynamic scenarios. These parameters may vary from situation-to-situation and person-to-person. By adjusting the parameters, an agent could adapt to each individual.

5.2 Environmental Interference

The study conducted in this work was performed in a controlled idealized social scenario; however, real-world scenarios will include multiple sources of environmental interference, including noise from a crowded room and/or low-light conditions [18]. An autonomous agent might encounter such dynamic and unstructured factors in social encounters; thus, future work will examine the use of the system in these real-world scenarios, examining its ability to recognize the metrics described, as well as exploring additional techniques to handle potential errors. In particular, we will focus on the impact of such interference on Hall's metrics [35].

5.3 Human-Robot Interaction

As autonomous intelligent robots become more ubiquitous in society, it will become necessary for them to interact with humans using natural modalities. Recognition of spatial behaviors, such as those described in this work, is a fundamental component for these agents to interact in an appropriate manner. By combining this study of proxemics with techniques from robotics and machine learning, we can create robust, socially-aware robots. Sensor data fed into a predictive model would allow the agent to recognize high-level interaction cues. A generative model incorporated into an action selection system would allow the agent to produce the similar interaction cues as it was trained to recognize, in an effort to produce a predictable social response from the human.

Acknowledgements. This work is supported in part by an NSF Graduate Research Fellowship, as well as ONR MURI N00014-09-1-1031 and NSF CNS-0709296 grants, and the Dan Marino Foundation. We would like to thank Louis-Philippe Morency for his insight in the experimental design process, and Mark Bolas and Evan Suma for their assistance in using the PrimeSensor. We would also like to thank Edward Kaszubski for his help in constructing the physical setup, and Dave Feil-Seifer for his assistance in using the overhead camera and Overhead Interaction Toolkit.

References

1. Hall, E.T.: The Silent Language. Doubleday Co., New York (1959)
2. Hall, E.T.: Handbook for Proxemic Research. American Anthropology Assn., Washington (1974)
3. Deutsch, R.D.: Spatial Structurings in Everyday Face-to-face Behavior: A Neurocybernetic Model. In: The Association for the Study of Man-Environment Relations, Orangeburg (1977)
4. Mead, R., Atrash, A., Matarić, M.J.: Recognition of Spatial Dynamics for Predicting Social Interaction. In: HRI 2011, Lausanne (2011)
5. Mehrabian, A.: Nonverbal Communication. Aldine Transaction, Piscataway (1972)
6. McNeill, D.: Gesture and Thought. Chicago University Press, Chicago (2005)
7. Hall, E.T.: The Hidden Dimension. Doubleday Co., New York (1966)
8. Argyle, M., Dean, J.: Eye-Contact, Distance, and Affiliation. Sociometry 28, 289–304 (1965)
9. Burgoon, J.K., Stern, L.A., Dillman, L.: Interpersonal Adaptation: Dyadic Interaction Patterns. Cambridge University Press, New York (1995)
10. Schöne, H.: Spatial Orientation: The Spatial Control of Behavior in Animals and Man. Princeton University Press, Princeton (1984)
11. Hayduk, L.A., Mainprize, S.: Personal Space of the Blind. Social Psychology Quarterly 43(2), 216–223 (1980)
12. Kennedy, D.P., Gläscher, J., Tyszka, J.M., Adolphs, R.: Personal Space Regulation by the Human Amygdala. Nature Neuroscience 12, 1226–1227 (2009)
13. Kendon, A.: Conducting Interaction - Patterns of Behavior in Focused Encounters. Cambridge University Press, New York (1990)
14. Price, G.H., Dabbs Jr., J.M.: Sex, Setting, and Personal Space: Changes as Children Grow Older. Personal Social Psychology Bulletin 1, 362–363 (1974)
15. Aiello, J.R., Aiello, T.D.: The Development of Personal Space: Proxemic Behavior of Children 6 through 16. Human Ecology 2(3), 177–189 (1974)
16. Jones, S.E., Aiello, J.R.: Proxemic Behavior of Black and White First-, Third-, and Fifth-Grade Children. Journal of Personality and Social Psychology 25(1), 21–27 (1973)
17. Aiello, J.R.: Human Spatial Behavior. In: Handbook of Environmental Psychology, ch. 12. John Wiley & Sons, New York (1987)
18. Adams, L., Zuckerman, D.: The Effect of Lighting Conditions on Personal Space Requirements. The Journal of General Psychology 118(4), 335–340 (1991)
19. Geden, E.A., Begeman, A.V.: Personal Space Preferences of Hospitalized Adults. Research in Nursing and Health 4, 237–241 (1981)
20. Evans, G.W., Wener, R.E.: Crowding and Personal Space Invasion on the Train: Please Don't Make Me Sit in the Middle. Journal of Environmental Psychology 27, 90–94 (2007)

21. Aiello, J.R., Thompson, D.E., Brodzinsky, D.M.: How Funny is Crowding Anyway? Effects of Group Size, Room Size, and the Introduction of Humor. Basic and Applied Social Psychology 4(2), 192–207 (1981)
22. Cassell, J., Sullivan, J., Prevost, S., Churchill, E.F.: Embodied Conversational Agents. MIT Press, Cambridge (2000)
23. Kuzuoka, H., Suzuki, Y., Yamashita, J., Yamazaki, K.: Reconfiguring Spatial Formation Arrangement by Robot Body Orientation. In: HRI 2010, Osaka, pp. 285–292 (2010)
24. Walters, M.L., Dautenhahn, K., Boekhorst, R.T., Koay, K.L., Syrdal, D.S., Nehaniv, C.L.: An Empirical Framework for Human-Robot Proxemics. In: New Frontiers in Human-Robot Interaction, Edinburgh, pp. 144–149 (2009)
25. Huettenrauch, H., Eklundh, K.S., Green, A., Topp, E.A.: Investigating Spatial Relationships in Human-Robot Interaction. In: IROS 2006, Beijing (2006)
26. Mumm, J., Mutlu, B.: Human-Robot Proxemics: Physical and Psychological Distancing in Human-Robot Interaction. In: HRI 2011, Lausanne, pp. 331–338 (2011)
27. Takayama, L., Pantofaru, C.: Influences on Proxemic Behaviors in Human-Robot Interaction. In: IROS 2009, St. Louis (2009)
28. Bailenson, J.N., Blascovich, J., Beall, A.C., Loomis, J.M.: Equilibrium Theory Revisited: Mutual Gaze and Personal Space in Virtual Environments. Presence 10(6), 583–598 (2001)
29. Trautman, P., Krause, A.: Unfreezing the Robot: Navigation in Dense, Interacting Crowds. In: IROS 2010, Taipei (2010)
30. Feil-Seifer, D.J., Matarić, M.J.: Automated Detection and Classification of Positive vs. Negative Robot Interactions with Children with Autism using Distance-based Features. In: HRI 2011, Lausanne, pp. 323–330 (2011)
31. Oosterhout, T.V., Visser, A.: A Visual Method for Robot Proxemics Measurements. In: HRI 2008 Workshop on Metrics for Human-Robot Interaction, Technical Report 471, Amsterdam (2008)
32. Jones, S.E., Aiello, J.R.: A Test of the Validity of Projective and Quasi-projective measures of Interpersonal Distance. The Western Journal of Speech Communication 43, 143–152 (1979)
33. Schegloff, E.A.: Body Torque. Social Research 65(3), 535–596 (1998)
34. McNeill, D.: Gesture, Gaze, and Ground. LNCS, pp. 1–14. Springer, Heidelberg (2005)
35. Hall, E.T.: A System for Notation of Proxemic Behavior. American Anthropologist 65, 1003–1026 (1963)
36. Sommer, R.: Sociofugal Space. The American Journal of Sociology 72(6), 654–660 (1967)
37. Lawson, B.: Sociofugal and Sociopetal Space, The Language of Space. Architectural Press, Oxford (2001)
38. Low, S.M., Lawrence-Zúñiga, D.: The Anthropology of Space and Place: Locating Culture. Blackwell Publishing, Oxford (2003)
39. Mead, R., Matarić, M.J.: An Experimental Design for Studying Proxemic Behavior in Human-Robot Interaction. Technical Report CRES-11-001, USC Interaction Lab, Los Angeles (2011)
40. Mead, R., Atrash, A., Matarić, M.J.: A Preliminary Validation of Automated Feature Recognition for Proxemic Behavior Analysis. Technical Report CRES-11-002, USC Interaction Lab, Los Angeles (2011)
41. Morency, L.P., Sidner, C., Lee, C., Darrell, T.: Contextual Recognition of Head Gestures. In: International Conference on Multimodal Interactions, Torento, pp. 18–24 (2005)

Children Interpretation of Emotional Body Language Displayed by a Robot

Aryel Beck[1], Lola Cañamero[1], Luisa Damiano[1],
Giacomo Sommavilla[2], Fabio Tesser[2], and Piero Cosi[2]

[1] Adaptive Systems Research Group, School of Computer Science & STRI,
University of Hertfordshire, United Kingdom
{a.beck2,l.canamero,l.damiano}@herts.ac.uk
[2] Institute of Cognitive Sciences and Technologies, Padova, Italy
{giacomo.sommavilla,fabio.tesser,piero.cosi}@pd.istc.cnr.it

Abstract. Previous results show that adults are able to interpret different key poses displayed by the robot and also that changing the head position affects the expressiveness of the key poses in a consistent way. Moving the head down leads to decreased arousal (the level of energy), valence (positive or negative) and stance (approaching or avoiding) whereas moving the head up produces an increase along these dimensions [1]. Hence, changing the head position during an interaction should send intuitive signals which could be used during an interaction. The ALIZ-E target group are children between the age of 8 and 11. Existing results suggest that they would be able to interpret human emotional body language [2, 3].

Based on these results, an experiment was conducted to test whether the results of [1] can be applied to children. If yes body postures and head position could be used to convey emotions during an interaction.

1 Introduction

The ALIZ-E project aims to contribute to the development of integrated cognitive systems capable of naturally interacting with young people in real-world situations, with a specific goal of supporting children engaged in a residential diabetes-management course. Fundamental to making human-robot interaction natural and integrated into the fabric of our lives, is that the robot can establish itself cognitively in the long term. Only if interaction provides a sense of continuity over longer periods of time, can it provide the resonance necessary for a constructive relationship between human and robot. It is commonly acknowledged that learning, adaptation, emotion, multi-modal dyadic and group interactions will be necessary to achieve this goal, but the field has not yet been presented with conclusive design paradigms, algorithms and test results showing how a robot can enter and successfully maintain an interaction spread beyond the current single episode interaction frame and stretching over several days. The work reported in this paper focuses on the emotional aspect. More precisely, it is concerned with developing methods that will enable a robot to display emotions in a way that can be readily interpreted by children during an interaction.

B. Mutlu et al. (Eds.): ICSR 2011, LNAI 7072, pp. 62–70, 2011.

Existing work in achieving expressive agents is difficult to apply to humanoid robots such as Nao. Work has been conducted on computer agents that achieve responsive behaviors using bodily expressions. For instance, Gillies et al. (2008) have designed a method to create responsive virtual humans that can generate expressive body language while listening to a human. Their expressions are based on motion capture data [4]. However, it would be difficult and tedious to transfer this method onto robots directly as they cannot reproduce the movements recorded by motion capture as smoothly as virtual humans or without causing the robot to often lose its balance. Expressive robots have also been successfully created. For instance, Kismet expresses emotions through its face [5]. Its expressions are based on nine prototypical facial expressions that 'blend' (interpolate) together along three axes: Arousal, Valence and Stance. Arousal is defined as the level of energy. Valence specifies how positive or negative the stimulus is. Stance reflects how approachable the stimulus is. This method defines an Affect Space, in which expressive behaviours span continuously across these three dimensions, allowing a wide range of expressions [5]. The method is interesting for its simplicity; however, the stance dimension may be problematic as it is not related to any accepted model of emotions. Modeling expressive emotions based on notion that are not validated in psychology may be problematic for long term interaction outside the laboratory. It may result in displaying inappropriate behaviour which could be counterproductive for the interaction. Moreover, for many robots such as Nao the same Affect Space cannot be directly applied as they do not have the ability to display facial expressions. The only medium available for such robots to express emotions is their bodies and voices.

2 Body Language as a Modality for Robot to Display Emotions

It has been shown that body language can be interpreted accurately without facial or vocal cues [6-8]. These results suggest that a humanoid robot, such as Nao, should be able to display emotions using its body. This is further suggested by traditional animation which focuses on the display of emotion through the body in order to increase believability. It has been codified as a rule in classical animations: "the expression must be captured throughout the whole body as well as in the face" [9]. Theatre follows a similar principle, by asking actors to become, in Artaud's words, "athletes of the emotions". Moreover, a large part of an actor's training addresses the non-verbal expression of emotions. These suggest that emotions such as fear, anger, happiness, stress, etc., could be readable when expressed through the body of Nao. However, most of the research on the psychology of emotions has focused on facial expressions. This makes it difficult to create an expressive systems based on bodily expressions. There is no equivalent to the Facial Action Coding System [10] for body expressions. Researchers have categorized the different types of body language, depending on how they occur. The categorization presented below was created from [11, 12] and classifies body language into three different areas.

Postures: Postures are specific positioning that the body takes during a timeframe. It has been established that postures are an effective medium to express emotion. For instance, De Silva et al. (2004) investigated cross-cultural recognition of four emotions (anger, fear, happiness, sadness) through interpretations of body postures. They built a set using actors to perform emotional postures and showed that it was possible for participants to correctly identify the different emotions [13]. Thus, a humanoid robot displaying emotion should take up postures appropriate to the emotion.

Movement: It has been shown that many emotions are differentiated by characteristic body movements, and that these are effective cues for judging the emotional state of other people even in the absence of facial and vocal cues [14]. Thus, a Nao robot displaying emotion should also do so during, and via, motion. Body movements include the movements themselves as well as the manner in which they are performed (i.e. movement speed, dynamics, curvature, etc.). The movements' dynamics have been shown to contribute to the emotional expression. For instance, Wallbott (1998) compared body language displayed by actors portraying different emotional states and found significant differences in the movement dynamics as well as in the type of movements performed across emotions [15]. Pollick et al (2001) investigated affect from point-light display of arm movements and found that activation is a formless cue that relates directly to the kinematics of the movements [16]. These studies are interesting because they show that dynamics is an essential component of an emotional expression.

Proxemics: It is the distance between individual during a social interaction. It is also indicative of emotional state. For example, angry people have a tendency to reduce the distance during social interaction. The problem is that this would also be the case between intimate people. Hence, proxemics cannot therefore be considered as an emotional expression in itself but is required to complete a representation of emotional behaviour and could be an interesting addition for the expressivity of a robot.

3 Previous Results

In animation, one of the established methods for creating convincing and believable displays consists in starting from the creation of expressive key poses (i.e. postures) rather than body language in motion [9]. In the context of emotional body language, a key pose is a static posture modelled so that it clearly describes the emotion displayed. Once the key poses are realized in robotic platforms, they can be used to drive the expressive animated behaviours. This method of creation was selected for the robot because it is possible to independently manipulate the position of joints and test the effects on the expressiveness of the key poses. If expressive key poses can be automatically generated by changing the position of a subset of joints, they can then be used to drive the expressive behaviours of the robot.

Previous work focused on validating a set of key poses and on testing the effect of moving the head up or down in a range of different key poses [1]. The position of the head was chosen because of its importance regarding the expression of emotions [17, 18]. Moreover, animation emphasizes the importance of creating strong silhouette [18, 19] and it is expected that manipulating the head position will considerably change a robot's silhouette.

This experiment showed that it was possible for adults to interpret the different key poses displayed by the robot and also that changing the head position affects the expressiveness of the key poses in a consistent way. It was found that moving the head down leads to decreased arousal (defines the level of energy), valence (defines whether a stimulus is positive or negative) and stance (defines whether a stimulus is approachable) whereas moving the head up increases these three dimensions [1]. This suggests that changing the head position during an interaction should send intuitive signals which will be used, for example, to indicate whether an interaction is successful.

These results were established with adults. However, the ALIZ-E project focuses on children and it is therefore necessary to test whether they can be extended to such a specific population. The results could depend on cultural and age differences.

4 Research Question

According to Boone and Cunningham's research on developmental acquisition of emotion decoding from expressive body movement [2, 3], as children begin to produce certain actions, they have access to the perceptual expressive cues associated to these actions. In turn, this can lead to effective cue utilisation. Boone and Cunningham experiment shows that, with respect to adults, it is possible to associate cues in naturally generated dance expression to specific emotions, and that children, from 8 years of age, can recognise them for the target emotions of happiness, sadness, anger, and fear [3]. However, existing studies have also shown that emotional recognition continues to develop during the adolescence [20]. Additionally, research in the perception of robots, suggests that there may be differences in the way children and adults perceive them [21]. It is therefore not evident that children and adults would interpret the body language displayed by a robot similarly. Thus, the purpose of the study reported in this paper was to test the results of [1] with children and to investigate whether the head position could be used to convey different emotions to such a specific population.

5 The Experiment

The experiment setting was defined to be as similar as possible to the one used with adult participants [1]. It used a within-subjects design with two independent variables: *Emotion Displayed* and *Head Position*. The effect of changing the head position may vary depending on the position of other joints. In other words, the effect of moving the head up or down may differ depending on the emotion being displayed. Therefore, it was tested across six emotions (*Emotion Displayed*): Anger, Sadness, Fear, Pride, Happiness and Excitement (Table 1).

Head position had three levels (Up, Down, and Straight), defined as the head position relative to the chest. One dependent variable was defined to explore the Affect Space: *Correct Identification. It* was used to test whether or not it was possible for participants to interpret the emotion of the key poses. Although the study conducted on adults was investigating *Arousal*, *Valence* and *Stance* as well, it was decided to remove them from this study because of the age difference.

The three main questions tested were:

(Q1) Are children as accurate as adults in identifying the key poses displayed by Nao?

(Q2) What is the effect of changing the head position on the interpretation and perceived place of a key pose in the Affect Space?

(Q3) Is the effect of moving the head similar across all the key poses? In other words, is the contribution of head position independent from the rest of the expression?

5.1 Participants

24 Children (13 females, 11 males) were recruited from the school "scuola media Dante Alighieri" (Italy) ranging in age from 11 to 13 (M=12, SD=0.3).

5.2 Material

The same material as in the study conducted with adults was reused. The reader can refer to [1] for a detailed report on the construction of the material. The platform chosen for this experiment was Nao, a humanoid robot with 25 degrees of freedom.

The experimental poses were generated by systematically altering the head positions of 6 emotional key poses. For *Head Position-Down,* the head was rotated vertically all the way down. For *Head Position -Up*, the head was moved vertically completely up. For *Head Position-straight*, the head was aligned with the chest. This resulted in 18 poses (6 *Emotion Displayed* by 3 *Head Positions*).

5.3 Procedure

The same experimenters tested all participants in groups of four. Participants were given full explanation regarding the questionnaire that they were expected to answer and were instructed to *"imagine that the robot is reacting to something"*. After confirming that they understood all the questions, participants watched and assessed the 18 poses. Each pose was displayed only once in a randomized order different for each group of participants. For each pose, participants were asked to assign an emotion label chosen from a list of six emotions. The list was comprised of Anger, Sadness, Fear, Pride, Happiness and Excitement. When all the poses were assessed, participants were fully debriefed. The sessions lasted approximately 30 minutes.

6 Results

6.1 Identification of the Emotion

Table 1. Percentage of participants who correctly identified the emotional key pose at least once (Chance level would be 42%)

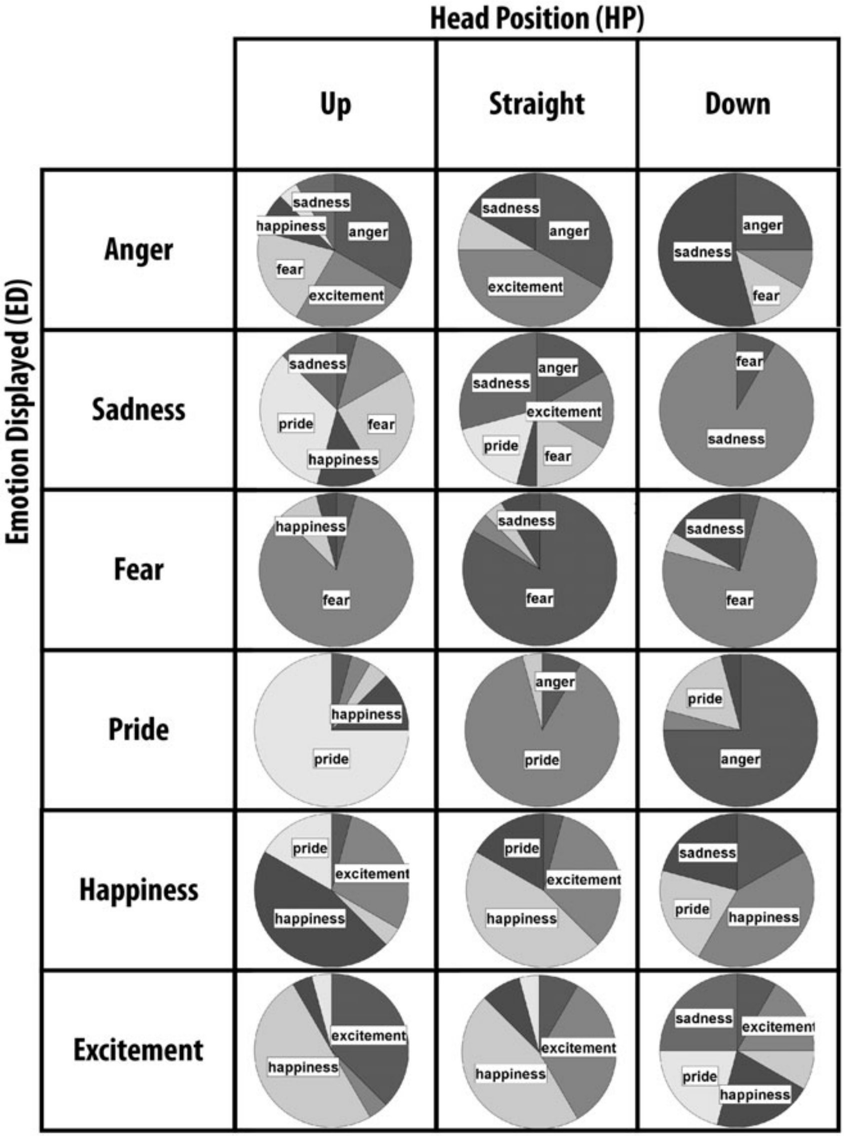
Pride	Happiness	Excitement	Fear	Sadness	Anger
100%	83%	63%	92%	92%	58%

Repeated Measures ANOVA (6 *Emotion Displayed* x 3 *Head Position)* was conducted on *Correct Identification. Emotion Displayed* had a significant effect on *Correct Identification (CI)* (F(5,115)=12.03, p<0.01, Partial η^2=0.34). *Head Position* had no significant main effect on *Correct Identification* (F(2,46)=1.45, p=0.25, Partial η^2=0.06).

These results indicate that participants' performance was different across emotions. Table 1 shows that the children correctly identified each emotion from viewing only the key pose of Nao. Recognition rates were above chance level although they varied from 58% for anger, to 100% for pride. Chance level would be $(1\text{-}(5/6)^3)*100= 42\%$.

6.2 Effect of Head Position on the Interpretation

Table 2. Effect of Head Position on the Interpretations of the body language displayed



There was a significant interaction between *Emotion Displayed* and *Head Position* ($F(10,230)=9.32$, $p<0.01$, Partial $\eta^2=0.29$). This indicates that the effect of *Head Position* on *Correct Identification* depended on the individual emotion being displayed. Therefore, the effect of *Head Position* were considered separately for each emotion and are reported in Table 2. It shows how the emotional interpretations of the displays shifted as a function of both the *Emotion Displayed* and the *Head Position*. The patterns found in this study are comparable to the one found with adult participants [1]. Participants were better at interpreting the negative emotions when the *Head Position* was Straight or Down. Participants were better at interpreting the positive emotions when the *Head Position* was Up.

7 Discussions

The first goal of the study was to test the expressivity of the key poses displayed by the robot with children. As with adults, the results show that the children who participated in the study were far better than chance level at interpreting the different key poses taken by the robot (Table 1). These recognition rates were obtained using static key poses only. Moreover, the relatively low recognition rates for Happiness and Excitement were mainly due to these two emotions being mistaken for one another (Table 2). These results clearly show that it is possible for children to interpret emotions displayed by a humanoid robot and that the lack of facial expression is not a barrier to expressing emotions. This suggests that they could be used to improve robots social skills. This is important as social robots need to be able to express their internal states in order to interact with humans in a natural and intuitive way.

As in [1], Head Position had a strong effect on the interpretation of the key poses being displayed (Table 2). For instance, children's interpretations of the Pride display were very similar to those of the adults. More precisely, it was interpreted as Pride when the head was up or straight. However, with the head down, a majority of children interpreted it as anger (Table 2). Fear was not affected by the change in Head Position and was correctly interpreted in all conditions both by the adults and the children. This further suggests that the interpretation of the key poses was similar in the adult and children's testing conditions.

Interestingly, children were less accurate than adults at interpreting the 'anger' key pose (58% vs. 89%). This difference could be due to cultural or age differences or to the different settings between the two experiments. This is an interesting issue and should be explored in future research as it is not possible to draw definitive conclusions from this study. Moreover, it is important to highlight that the material used for this study is prototypical and was intentionally selected to be expressive. This is appropriate within the ALIZ-E project; however, it is likely that the use of prototypical expressions had an effect on the results and on the similarities of the interpretations that were found in this study.

8 Conclusion

As with adults, it was found that moving the head up increased the identification of some emotions (pride, happiness, and excitement), whereas moving the head down increased correct identification for other displays (anger, sadness). Fear, however, was well identified regardless of Head Position.

This has design implication for improving emotional body language displayed by robots. The results of this study suggest that the expressivity of the negative emotions (anger and sadness) can be improved by moving the head down while the expressivity of the positive emotion (happiness, excitement and pride) can be improved by moving the head up. These results have already been successfully integrated in an automated expressive system [22]. The robot can automatically change its head position to express changes in its internal state.

Future work will explore the effect of moving the different parts of the body on the interpretation of the body language displayed as well as adding dynamic elements to the expressions. If similar results can be established for the other parts of the body, it will be possible to create a rich Affect Space for humanoid robots.

Acknowledgments. The authors would like to thank the school "scuola media Dante Alighieri" for hosting the experiment. This work is funded by the EU FP7 ALIZ-E project (grant number 248116).

References

1. Beck, A., Cañamero, L., Bard, K.: Towards an affect space for robots to display emotional body language. In: Ro-Man 2010. IEEE, Viareggio (2010)
2. Boone, R.T., Cunningham, J.G.: The Attribution of Emotion to Expressive Body Movements: A Structural Cue Analysis (1996) (manuscript)
3. Boone, R.T., Cunningham, J.G.: Children's decoding of emotion in expressive body movement: the development of cue attunement. Dev. Psychol. 34(5), 1007–1016 (1998)
4. Gillies, M., et al.: Responsive listening behavior. Computer Animation and Virtual Worlds 19(5), 579–589 (2008)
5. Breazeal, C.: Designing sociable robots. In: Intelligent Robotics & Autonomous Agents. MIT Press, Cambridge (2002)
6. Beck, A., Stevens, B., Bard, K.: Comparing perception of affective body movements displayed by actors and animated characters. In: AISB 2009, Edinburgh, UK (2009)
7. de Gelder, B.: Towards the neurobiology of emotional body language. Nature Reviews Neuro-Science 7(3), 242–249 (2006)
8. Kleinsmith, A., De Silva, P.R., Bianchi-Berthouze, N.: Cross-cultural differences in recognizing affect from body posture. Interacting with Computers 18(6), 1371–1389 (2006)
9. Thomas, F., Johnston, O.: The illusion of life. Abbeville Press, New-York (1995)
10. Ekman, P., Friesen, W.V., Hager, J.C.: Facial action coding system. The manual. Human Face, Salt Lake (2002)
11. Cassell, J.: Nudge nudge wink wink: elements of face-to-face conversation. In: Cassell, J., et al. (eds.) Embodied Conversational Agents, pp. 1–27. MIT Press, Cambridge (2000)

12. Vinayagamoorthy, V., et al.: Building Expression into Virtual Characters. In: Eurographics 2006. Proc. Eurographics, Vienna (2006)
13. De Silva, P.R., Bianchi-Berthouze, N.: Modeling human affective postures: an information theoretic characterization of posture features. Computer Animation and Virtual Worlds 15(3-4), 269–276 (2004)
14. Atkinson, A.P., et al.: Emotion perception from dynamic and static body expressions in point-light and full-light displays. Perception 33, 717–746 (2004)
15. Wallbott, H.G.: Bodily expression of emotion. European Journal of Social Psychology 28(6), 879–896 (1998)
16. Pollick, F.E., et al.: Perceiving affect from arm movement. Cognition 82(2), 51–61 (2001)
17. Kleinsmith, A., Bianchi-Berthouze, N., Steed, A.: Automatic Recognition of Non-Acted Affective Postures. IEEE Trans. Syst. Man Cybern. B Cybern. (2011)
18. Schouwstra, S., Hoogstraten, J.: Head position and spinal position as determinants of perceived emotional state. Perceptual and Motor Skills 81(2), 673–674 (1995)
19. Maestri, G.: Digital character animation. In: Kissane, E., Kalning, K. (eds.), 3rd edn. New Riders, Berkeley (2006)
20. Tonks, J., et al.: Assessing emotion recognition in 9-15-years olds: preliminary analysis of abilities in reading emotion from faces, voices and eyes. Brain Inj. 21(6), 623–629 (2007)
21. Woods, S., Dautenhahn, K., Schultz, J.: Child and adults' perspectives on robot appearance. In: AISB 2005 Symposium on Robot Companion, pp. 126–132. SSAISB, Hatfield (2005)
22. Beck, A., et al.: Interpretation of Emotional Body Language Displayed by Robots. In: Affine 2010. ACM, Firenze (2010)

Making Robots Persuasive: The Influence of Combining Persuasive Strategies (Gazing and Gestures) by a Storytelling Robot on Its Persuasive Power

Jaap Ham[1], René Bokhorst[1], Raymond Cuijpers[1],
David van der Pol[1], and John-John Cabibihan[2]

[1] Human-Technology Interaction Group,
Department of Industrial Engineering and Innovation Sciences,
Eindhoven University of Technology, Eindhoven, The Netherlands
j.r.c.ham@tue.nl
[2] Social Robotics Laboratory (Interactive and Digital Media Institute)
& Department of Electrical and Computer Engineering,
National University of Singapore, Singapore
elecjj@nus.edu.sg

Abstract. Social agency theory suggests that when an (artificial) agent combines persuasive strategies, its persuasive power increases. Therefore, we investigated whether a robot that uses two persuasive strategies is more persuasive than a robot that uses only one. Because in human face-to-face persuasion two crucial persuasive strategies are gazing and gestures, the current research investigated the combined and individual contribution of gestures and gazing on the persuasiveness of a storytelling robot. A robot told a persuasive story about the aversive consequences of lying to 48 participants. The robot used persuasive gestures (or not) and gazing (or not) to accompany this persuasive story. We assessed persuasiveness by asking participants to evaluate the lying individual in the story told by the robot. Results indicated that only gazing independently led to increased persuasiveness. Using persuasive gestures only led to increased persuasiveness when the robot combined it with (the persuasive strategy of) gazing. Without gazing, using persuasive gestures diminished robot persuasiveness. The implications of the current findings for theory and design of persuasive robots are discussed.

Keywords: Social Robotics, Persuasive Robotics, Gazing, Head Movement, Gestures, Storytelling Robot, Nao, Persuasion, Persuasive Technology.

1 Introduction

The development of robotics in the past few years has shown that robots are capable of doing many things. For example, robots assist in medical procedures [1], and robots are deployed for dangerous tasks [2]. Beginning as a discipline within electrical engineering, the research field of robotics has expanded into the field of psychology because of the increasing frequency of encounters between humans and robots. We

B. Mutlu et al. (Eds.): ICSR 2011, LNAI 7072, pp. 71–83, 2011.

argue that one of the biggest challenges in robotics in the coming decades is to engineer robots that can effectively influence human behavior or attitudes in the sense that they facilitate humans, and are able to effectively interact with humans. Whether it is actual behavior change (e.g., help the human walk better, or take her pills), or a more cognitive effect like attitude change (inform the human about danger), or even changes in cognitive processing (help the human learn better), effective influencing humans is fundamental to developing successful social robots.

Therefore, we argue that for the development of social robots, knowledge about how to make these social robots persuasive is crucial to be able to construct robots that can effectively help human users. So, it is important to develop the research field of *persuasive robotics*: the scientific study of artificial, embodied agents (robots) that are intentionally designed to change a person's behavior, attitudes, and/ or cognitive processes [18, see also, 19, 20, 21].

Social agency theory [12] suggests that when an (artificial) agent combines persuasive strategies, its persuasive power increases. That is, social agency theory suggests that social cues can prime a social conversation schema in people, and cause people to act as if they were in a conversation with another person [12], which is in line with the Media Equation theory [5]. Based on these theories, in general, more persuasive strategies should lead to more persuasion. Therefore, we investigated whether a robot that uses two persuasive strategies is more persuasive than a robot that uses only one.

In persuasion by humans, two of the most crucial elements of face-to-face communication that determine persuasion (through steering attention and comprehensibility) are gazing [3] and gestures [4]. As media equation research suggests [5], in many forms of human-computer communication as well as human-robot communication humans respond and interact with technology (computers or robots) just as if they were interacting with other humans. Therefore, we argue that the importance of gazing behavior and use of gestures for persuasion by robots will also be very high. Therefore, the current research investigated the combined and individual contribution of gestures and gazing on the persuasiveness of a storytelling robot.

Earlier research suggested that a robot's *gazing* behavior (looking at a participants' face) can influence its persuasiveness and various variables that are related to persuasion. For example, research [22] suggested that a robot that tracks a user (moving its head or eyes such that they stay in the direction of the user) is more persuasive than a robot that does not track a user. Also, research indicated that a robot's gazing behavior influences information retention by their human interaction partners [6]. That is, participants who were gazed at more, significantly remember more of a story the robot told than people who were gazed at less. Also, other earlier research [7] suggests that the feeling of involvement in a conversation is for a large part dependent on the ability to see what our (artificial) conversation partner is looking at. Likewise, an earlier study [8] showed that participants controlling an avatar in a virtual environment who can move their virtual heads, will like other avatars controlled by participants more and experience a higher level of co-presence compared to conditions where no gazing or head movement was possible. In a different study [9] it was found that a person's gaze is a very reliable source of input to establish whom the person is speaking or listening to. Head movement and gazing are shown to be very important factors in co-presence, likeability and determining direction [9]. It was also found [3] that gazing can play an important role in compliance to unambiguous requests in human-human interaction.

Earlier research also suggested that a robot's *gesture* behavior can influence various variables that are related to persuasion, although direct evidence for the influence of robot gestures for its persuasiveness is still lacking. For example, earlier research [10] suggested that pointing gestures by a robot can significantly improve the speech comprehension of spatial information spoken by a human experimenter. Also, earlier research suggested [23] positive effects for object identification by human users when a robot uses pointing gestures. Other research [11] suggested that, in general, gestures that are linked to speech influence message evaluation and judgments about the speaker more positively than gestures that are not linked to speech or speech without gestures. Furthermore, many studies of human-human communication suggested that gestures improve communication [4].

Therefore, we predicted that gazing and gestures would both significantly improve the persuasive power of a robot that delivers a persuasive message. The core question of the current research, however, is: Will a robot that uses two persuasive strategies (both gazing and gestures) be more persuasive than a robot that uses only one persuasive strategy (either gazing or gestures)? Thereby, the current research might (a) find evidence for the assumption of social agency theory [12] that persuasive strategies will have additive effects. Also, the current research might (b) replicate earlier research [e.g., 22] providing evidence for the persuasive power of robot gazing, and (c) provide evidence for the persuasive power of gestures used by a robot.

1.1 The Current Research

More specifically, the current research investigated whether a robot that tells a persuasive story would be more persuasive if it used gazing and gestures to accompany that story. In all four conditions of the experiment, the robot told the participant a classical-Greek persuasive story. In half the conditions the robot showed gestures that accompanied this story, whereas in the other half of the conditions it did not. Also, in half the conditions, the robot told the persuasive story while part of the time gazing at the participant, whereas in the other half of the conditions it did not. To construct the robot's gestures and gazing, we employed the methodology of earlier research [6], and videotaped a storyteller. This storyteller was asked to tell the same story (to a third person) as the one told by the robot in our experiment, and was asked to accompany telling that story with persuasive gestures. We programmed the robot's behavior such that it mimicked the actor's gazing behavior and gestures.

1.2 Hypotheses

Based on the research findings and theories described above, we had the following hypotheses.

H1: Participants who receive a persuasive message from a storytelling robot that employs human-like gazing behavior will be persuaded more.

H2: Participants who receive a persuasive message from a storytelling robot that employs human-like gestures will be persuaded more.
H3: The effect of adding either gazing behavior or gestures to the other will significantly increase persuasion.

2 Method

2.1 Participants and Design

Sixty-four participants (33 female and 31 male) participated in this experiment. Participants were students or researchers of the National University of Singapore or their friends or family. Their average age was between 13 and 32 years old (*M* = 22.50, *SD* = 2.63). For a participation of 20 minutes they received a compensation of $10 SGD.

For this experiment we employed a 2 (looking behavior: present vs. absent) x 2 (gestures: present vs. absent) between-participants design.

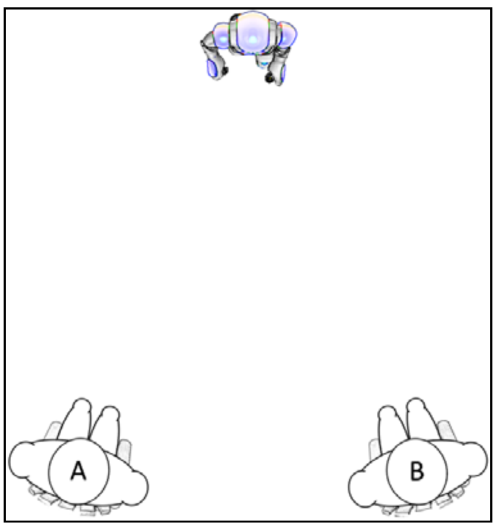


Fig. 1. Participants A and B listen to Nao's story as a pair

2.2 Materials and Measures

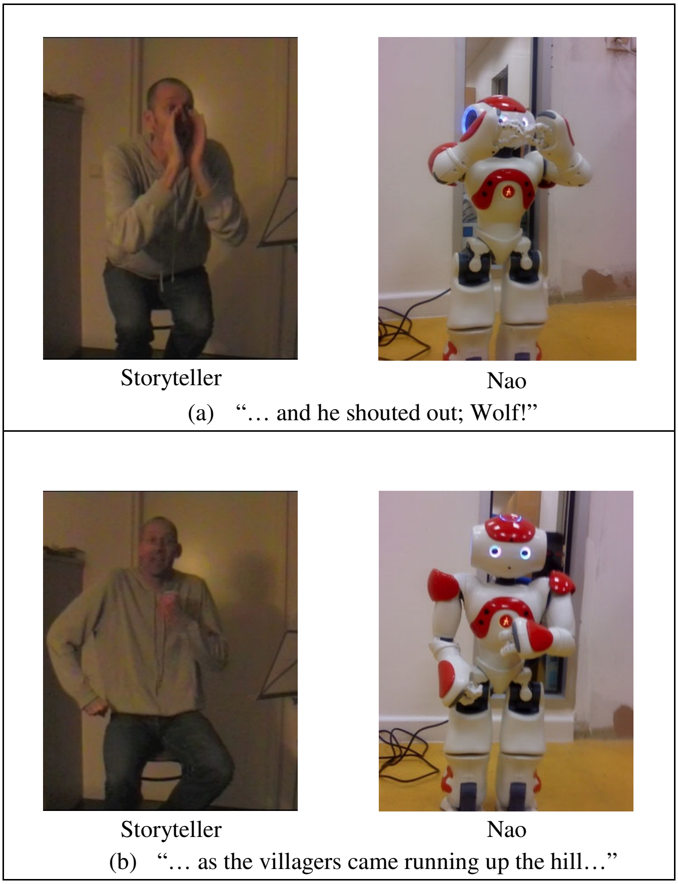
The persuasive story told to the participant was combined with gazing, gestures, both or no. The persuasive story was told by a Nao robot (see Figure 2). The persuasive story was the classical Greek story "The boy who cried wolf" by Aesop. This is a moral story to teach children about the (aversive) consequences of lying, and this story is quite unknown in Singapore (where the study was conducted).

Fig. 2. The Nao robot by Aldebaran

It should be noted that according to some definitions of gazing [7], [15], Nao is unable to gaze. That is, with only two sunken LED lights giving shape to its eyes, Nao is unable to show any kind of focus other than pointing the front of its head in a specific direction. We argue that even this minimal form of gazing can be argued to be gazing. However, to control for misinterpreted intentions of gazing we assessed using several questions posed at the end of the experiment how participants experienced Nao's gaze.

The movements (gestures and gazing) of the Nao robot were programmed based on a video recording of a professional stage actor telling the persuasive story to a third person being present and showing related, persuasive gestures. All the actor's motions were categorized (21 different gestures, 8 different gazing behaviors) and animated in Nao's special software (Choreographe). See Figure 3 for two examples of a gesture by the human actor, and the implementation into the Nao's behavior. Gestures consisted of shouting (raising hands to mouth, see Figure 3a), a simulation of walking behavior (see Figure 3b), various pointing gestures, et cetera. All behaviors were performed by the human actor while sitting on a high chair that did not impede leg movement, but that did keep the human actor at the same location. The total storytelling script that was constructed for Nao had a total running time of less than three minutes.

To assess persuasion, we assessed a participant's attitude about lying by asking them nine questions [see 16] in which they evaluated the lying individual (the shepherd boy in the story) after having listened to the storytelling robot. For all nine questions (each employing a different synonym), the participant rated his or her evaluation of the lying individual on a 7-point scale (-3 for negative, and +3 for positive). More specifically, participants were asked to rate his or her evaluation of the shepherd boy on a 7-point scale ranging from "bad" to "good", "negative" to "positive", "unfriendly" to "friendly", "dislikable" to "likable", "unpleasant" to "pleasant", "not nice" to "nice", "unagreeable" to "agreeable", "wrong" to "right", and "incorrect" to "correct". We averaged these nine answers to construct a reliable (alpha = .70) measure of persuasion, and the distribution of this measure was normal in all four conditions.



(a) "… and he shouted out; Wolf!"

(b) "… as the villagers came running up the hill…"

Fig. 3. Two examples (a and b) of gestures as performed by a storyteller and Nao

2.3 Procedure

Participants were invited to participate in a study on "human-robot interaction," and were invited through social network sites and pamphlets that were strategically placed on campus at the National University of Singapore. Participants were welcomed in the lobby of the building in which the experiment took place, and were then escorted to a separate meeting room on the first floor where they were briefed about the experiment. The actual experiment took place in the same meeting room. After the briefing the Nao robot was placed in the room at a set location. Participants were asked to remain seated in their chairs throughout the whole experiment. During one session, two participants participated at the same time, seated next to one another (see Figure 1). This way, the Nao could for a good reason avert its gaze from the participant, that is, to look at the other participant. Each participant was (unbeknownst to them) randomly placed in one of four conditions. The condition determined how many (if any) and which cues (gazing or gestures) were used by the Nao robot while it told the story. After a participant was seated, the Nao robot told them the persuasive story. After the robot had finished the story, participants answered the nine questions (evaluating the shepherd boy) assessing the extend to which they had been persuaded that lying is bad.

Next, the participant filled out the Godspeed questionnaire [14] to assess his or her overall evaluation of the Nao robot, and participants answered three additional questions about the robot and the story ("Did Nao ever look at you during the story?", "Did you perceive Nao as male or female?", "Did you understand the story?"). Finally, participants were thanked, paid en debriefed.

3 Results

A first analysis showed that 25% of participants in conditions in which Nao gazed at them did not indicate to have perceived that Nao had been gazing at them. The remaining 75% agreed that Nao had actually been gazing at them. Also, 25% of participants in conditions in which Nao never gazed at anyone claimed to have seen Nao gazing at them. Because for these participants our manipulation seems not to have been perceived as intended, the data from these two groups of 25% of our participants were excluded from further analyses. The remaining participants were 24 males and 24 females between the ages of 13 and 32 (M = 22.48, SD = 2.51). Importantly, analyses that used all 64 participants showed completely the same pattern of results. Participants' gender had no significant effects on persuasion or likeability, neither independently or in interaction with our two manipulations (gazing and gestures), all F's < 1..

3.1 Persuasion

To assess the amount of persuasion by the Nao robot, we analyzed participant's evaluation of the lying person in the story. Therefore, the average of the answers to the nine questions evaluating the lying person in the story was submitted to a 2 (gazing: absent vs. present) x 2 (gestures: absent vs. present) MANOVA, in which both factors were manipulated between participants. This analysis indicated that participants who were gazed at by the storytelling robot evaluated the lying person in the story more negatively (M = 0.26, SD = 0.59) than participants who were not gazed at by the storytelling robot (M = 0.61, SD = 0.61), $F(1, 44) = 5.07$, $p = 0.03$. In confirmation of the first hypothesis, this analysis thereby suggested that participants who were gazed at by the storytelling robot were persuaded more by the persuasive message of the robot (indicating that lying is wrong).

Furthermore, this analysis presented no evidence that participants who had been told this persuasive message by a robot that used gestures were persuaded more (and evaluated the lying person more negatively) than participants who had been told this persuasive message by a robot that did not use gestures, $F < 1$. This finding does not provide evidence to prove our second hypothesis.

Table 1. Participant's evaluation of the lying person in the story (the amount of persuasion) by gazing present or absent and gestures used or not used

		Gazing	
		Absent	Present
Gestures	Absent	0.27 $(0.29)_a$	0.45 $(0.59)_a$
	Present	0.95 $(0.66)_b$	0.07 $(0.56)_a$

Note: Means in rows and columns with different subscripts are significantly different at $p < 0.05$.

Finally, this analysis suggested an interaction between gestures and gazing, $F(1, 44) = 11.82$, $p = 0.001$. An overview of all means and standard deviations can be found in Table 1. More specifically, for participants who had been told this persuasive message by a robot that did not gaze at them, those participants who were told the persuasive message by a robot that did not use gestures evaluated the lying person more negatively ($M = 0.27$, $SD = 0.29$) than those participants who were told the persuasive message by a robot that did use gesture ($M = 0.95$, $SD = 0.66$), $F(1, 45) = 8.82$, $p = 0.01$. In contrast, results did not provide evidence that for participants who had been told this persuasive message by a robot that gazed at them, those participants who were told the persuasive message by a robot that did not use gestures evaluated the lying person more negatively ($M = 0.45$, $SD = 0.59$) than those participants who were told the persuasive message by a robot that did use gestures ($M = 0.07$, $SD = 0.56$), $F(1, 45) = 2.84$, $p = 0.10$.

Likewise, for participants who had been told this persuasive message by a robot that used gestures, those participants who were told the persuasive message by a robot that gazed at them evaluated the lying person more negatively ($M = 0.07$, $SD = 0.56$) than those participants who were told the persuasive message by a robot that did not gaze at them ($M = 0.95$, $SD = 0.66$), $F(1, 45) = 16.23$, $p < 0.001$. Again in contrast, results did not provide evidence that for participants who had been told this persuasive message by a robot that did not use gestures, those participants who were told the persuasive message by a robot that gazed at them evaluated the lying person more negatively ($M = 0.45$, $SD = 0.59$) than those participants who were told the persuasive message by a robot that did not gaze at them ($M = 0.27$, $SD = 0.29$), $F < 1$.

Finally, we checked for effects of our manipulations on the overall evaluation of the robot, by assessing the effects of our manipulation on the Godspeed questionnaire [14]. Therefore the average answers to the questions on each of the five godspeed dimensions were submitted to a 5 (godspeed dimension: anthropomorphism vs. animacy vs. likeability vs. perceived intelligence vs. perceived safety) x 2 (gazing: absent vs. present) x 2 (gestures: absent vs. present) MANOVA in which godspeed dimension was manipulated within participants. This analysis indicated no main effect of gazing $F < 1$, no main effect of gestures $F < 1$, nor an interaction of gazing x gestures, $F < 1$. This analysis indicated an (irrelevant) main effect of godspeed, $F(4, 176) = 95.70$, $p < 0.0001$, and also did not provide evidence for different effects of gazing or gestures on any of the godspeed dimensions, indicated by non-significant interactions of gazing x godspeed, $F(4, 176) = 1.25$, $p = 0.29$, and of gestures x godspeed, $F(4, 176) = 1.67$, $p = 0.16$. The three-way interaction of godspeed x gazing x gestures was also non-significant, $F < 1$. These latter findings suggested that the effects of our manipulations on persuasion were not related to the effects of our manipulations on a participant's evaluation of the robot.

4 Discussion

Based on social agency theory [12] and the Media Equation hypothesis [5], we argued that just as in human-human persuasion, gazing and gestures are crucial persuasive strategies that will also, additively, increase the persuasiveness of a robot. To find evidence for this assumption, the current research investigated the combined and individual contribution of robotic gestures and gazing on the persuasiveness of a storytelling robot. A robot told a persuasive story about the aversive consequences of lying to 48 participants. The robot used persuasive gestures (or not) and gazing (or not) to accompany this persuasive story. We assessed persuasiveness by asking participants to evaluate the lying individual in the story told by the robot. Results indicated that only gazing independently led to increased persuasiveness. Using persuasive gestures only led to increased persuasiveness when the robot used (the persuasive strategy of) gazing. Without gazing, using persuasive gestures diminished persuasiveness.

More specifically, our first hypothesis was that robot persuasiveness would increase when that robot accompanied its persuasive message by gazing. Results provided evidence for H1, that is, participants who were gazed at by the storytelling robot were persuaded more by the persuasive message of the robot than people who were not gazed at by the storytelling robot. This finding is in line with earlier findings of the persuasive effects of gazing in human-human persuasion [see 3].

Furthermore, our second hypothesis was that robot persuasiveness would increase when that robot accompanied its persuasive message with gestures. However, results did not provide evidence for H2. That is, overall, participants were not persuaded more by the robot that used gestures in addition to its persuasive message, than by the robot that did not. This finding does not replicate findings of earlier research in human-human persuasion, in which gestures were found to enhance persuasion [see e.g., 4]. There might be several explanations for this result. Among those might be the mechanical appearance and lack of fluidity of the motions of the robot. In line with this suggestion, additional analysis indicated that the use of gestures by the storytelling robot had no significant main effect on the animacy (one of the godspeed dimensions [14]) of the robot $F(1, 44) = 1.16$, $p = 0.23$. Therefore, it may have been the case that the gestures used by the robot were insufficient in fluidity or were not recognized correctly to lead to an increase in robot persuasive strength.

Importantly however, the current findings countered that explanation by showing that the effect of gestures was qualified by a statistically significant interaction of gestures x gazing. That is, in confirmation of our third hypothesis (H3), results provided evidence that the addition of gestures led to more persuasion only when the robot also accompanied its persuasive message with gazing. Without gazing, the addition of gestures led to a decrease of persuasion by the robot. This finding suggested that comparable to human-human persuasion, gestures can make a robot more persuasive, but only when the robot looks at the persuadee. Indeed, when the robot did not gaze at the participant, this person may have felt not being addressed by the robot which could have lead to the participant not recognizing itself as the recipient of the message and thereby the participant may not have activated relevant social conversation schemas sufficiently enough to be persuaded [12]. This conflict could lead to an interaction that would explain the absence of a main effect of gestures being detected. We argue that the use of gestures without gazing is so rare in social context that it might decrease the persuasive strength of the persuasive message. Indeed, a participant who was told the persuasive message by the robot that did not gaze at him or her but that did use persuasive gestures may have gotten the impression that the robot was telling its persuasive message to another person—the other participant in the experimental setup.

The latter finding also indicates that a robot that accompanied its persuasive message by both gestures and gazing was more persuasive than a robot that accompanied its persuasive message by only either gazing or gestures. This finding is in line with social agency theory [12] that argues that the more social cues an artificial agent employs, the more it will activate human social interaction schemata. Thereby, a robot that uses several social cues (i.e., gazing and gestures) will be able to be more persuasive than a robot that uses only one social cue (i.e., gazing or gestures only). This finding also fits to the Media Equation hypothesis [5] which would argue that just as in human-human persuasion employing several persuasive strategies will lead to an increase in persuasive power, because people interact with artificial agent in the same manner as when interacting with humans.

However, the finding that robot persuasive strategies can have additive effects on robot persuasive strength is different from findings of earlier research. That is, earlier research [13], suggested that the addition of either voice (using voice to utter a persuasive message instead of changes in lighting to convey that message) or the addition of embodiment (using an embodied robot as the source of that voice or lighting changes, versus using a computer case as the source) to persuasive messages both led to an increase of robot persuasive strength, but that the combination of both voice and embodiment did not lead to a further increase. A theoretical rationale for those findings was found in a slightly different explanation of the Media Equation hypothesis. That is, one could also argue based on that hypothesis, that only one social cue (either voice or embodiment) emitted by an artificial social agent is enough to trigger in the human user a complete set of responses developed for human-human interaction. Based on those findings and the current findings, we argue that future research might investigate the distinction between social cues (e.g., having voice, embodiment, gazing or gestures) and social persuasion strategies (using those social cues to persuade) and their additive effects on robot agency perceptions and robot persuasiveness.

As indicated, the current research first of all learns us that using one persuasive strategy (either gazing or gestures) on top of another (either gazing or gestures) does not simply lead to an addition of the persuasive power of each persuasive strategy on its own. Rather, what results indicated is that using (a specific set of) gestures without gazing deteriorates persuasiveness, whereas using the same (specific set of) gestures with gazing does not. This interaction compares the effects of the same specific gestures under different conditions of gazing, and the specific characteristics of these gestures are less relevant for that comparison. Therefore, we argue that the core finding of the current research might be generalizable over gestures with various characteristics. At the same time, the specific set of gestures used in the current research did not influence persuasiveness overall. Thereby, the current research did not replicate earlier research in human-human interaction in which gestures were found to enhance persuasion [see e.g., 4]. Future research mainly intended to provide evidence for that phenomenon might employ gestures that are more strongly persuasive, or have a stronger persuasive function in corroboration to a persuasive message.

Also, the persuasive story used in the current research (which was the same in all conditions) has shown to be suitable for finding the interaction between gestures and gazing. Thereby, the current research suggested that this effect might be generalizable over comparable other persuasive stories. At the same time, perhaps characteristics of the persuasive story may have diminished the persuasive (main) effects of the used gestures (and potentially also of the gazing used, although that manipulation did lead to a significant main effect). Future research mainly aimed at providing evidence for the persuasive effects of gestures might employ a persuasive story that is less strongly persuasive, such that persuasive effects of gestures can be more easily be effective.

5 Conclusion

So, will a robot that uses two persuasive strategies (both gazing and gestures) be more persuasive than a robot that uses only one persuasive strategy (either gazing or gestures)? The current research suggests a confirmation of the idea that also for artificial agents (a robot in this case) persuasive strategies combine and influence each other's persuasive power. However, importantly, the current research suggests that these processes do not only influence each other positively. That is, when a robot used gestures without gazing at the participant, its persuasiveness became less. Furthermore, the current research replicated earlier research providing evidence for the persuasive power of robot gazing [22]. This suggests that a robot can become more persuasive when it uses gazing to accompany its persuasive message, similar to research done in human-human interaction [3]. Finally, the current research did not provide evidence that a robot can become more persuasive when it uses persuasive gestures to accompany its persuasive message, which is a new finding in human-robot interaction research, but also similar to findings in human-human interaction [4]. However, current results provided evidence for this increase in persuasiveness only when the robot looked at the participant. When the robot did not look at the participant, the addition of gestures led to a decrease of robot persuasiveness. It would be interesting to investigate if people would be persuaded by a human storyteller that uses gestures but does not look at the person when telling a persuasive story, to compare those results with the results of the current experiment. We argue that the persuasive effect of gestures would greatly diminish also in human-human interaction, when the persuader does not look at the persuadee.

To conclude, the current research indicated that despite robots taking over many tasks of human beings already [1], [2], it may still be a bit too early to start deploying robots as social agents in social settings. The current research suggested that robots can become more persuasive when they use human-like gazing behavior or human-like persuasive gestures to accompany the persuasive messages that they utter. This study also suggested that adding multiple social cues can increase the persuasive power even more. Future research might investigate when adding human-like persuasive strategies lead to a social psychological version of uncanny valley effects [see 17], which might be different from uncanny valley effects caused by an increase of physical human-like characteristics of artificial agents. The current research suggests that there is a lot of persuasive potential in incorporating social cues and persuasive strategies when trying to improve the effectiveness of a robot.

References

1. Bloss, R.: Your next surgeon be a robot! Industrial Robot: An International Journal 38, 6–10 (2011)
2. Humphrey, C.M., Adams, J.A.: Robotic tasks for chemical, biological, radiological, nuclear and explosive incident response. Advanced Robotics 23, 1217–1232 (2009)
3. Kleinke, C.L.: Compliance to requests made by gazing and touching experimenters in field settings. Journal of Experimental Social Psychology 13, 218–223 (1977)
4. Hostetter, A.B.: When do gestures communicate? A meta-analysis. Psychological Bulletin 137, 297–315 (2011)
5. Reeves, B., Nass, C.: The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places (Center for the Study of Language and Information - Lecture Notes). Center for the Study of Language and Inf. (2003)
6. Mutlu, B., Forlizzi, J., Hodgins, J.: A storytelling robot: Modeling and evaluation of human-like gaze behavior. In: IEEE-RAS Conference on Humanoid Robots (Humanoids 2006), pp. 518–523 (2006)
7. Garau, M., Slater, M., Bee, S., Sasse, M.A.: The impact of eye gaze on communication using humanoid avatars. In: The CHI 2001 Conference on Human Factors in Computing Systems, pp. 309–316 (2001)
8. Bailenson, J.N., Beall, A.C., Blascovich, J.: Mutual gaze and task performance in shared virtual environments. The Journal of Visualization and Computer Animation 13, 313–320 (2002)
9. Vertegaal, R., Slagter, R., van der Veer, G., Nijholt, A.: There is more to Conversational Agents than meets the Eyes. In: The CHI 2001 Conference on Human Factors in Computing Systems, pp. 301–308 (2001)
10. Cabibihan, J.-J., So, W.C., Nazar, M., Ge, S.S.: Pointing Gestures for a Robot Mediated Communication Interface. In: Xie, M., Xiong, Y., Xiong, C., Liu, H., Hu, Z. (eds.) ICIRA 2009. LNCS, vol. 5928, pp. 67–77. Springer, Heidelberg (2009)
11. Maricchiolo, F., Gnisci, A., Bonaiuto, M., Ficca, G.: Effects of different types of hand gestures in persuasive speech on receivers' evaluations. In: Language and Cognitive Processes, pp. 239–266. Psychology Press (2009)
12. Mayer, R.E., Sobko, K., Mautone, P.D.: Social cues in multimedia learning: Role of speaker's voice. Journal of Educational Psychology 95, 419–425 (2003)
13. Vossen, S., Ham, J., Midden, C.: What Makes Social Feedback from a Robot Work? Disentangling the Effect of Speech, Physical Appearance and Evaluation. In: Ploug, T., Hasle, P., Oinas-Kukkonen, H. (eds.) PERSUASIVE 2010. LNCS, vol. 6137, pp. 52–57. Springer, Heidelberg (2010)

14. Bartneck, C., Kulic, D., Croft, E.: Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence and perceived safety of robots. International Journal of Social Robotics 1, 71–81 (2009)
15. Cassell, J., Pelachaud, C., Badler, N., Steedman, M., Achorn, B., Becket, T., Douville, B., Prevost, S., Stone, M.: Animated conversation: rule-based generation of facial expression, gesture & spoken intonation for multiple conversational agents. In: SIGGRAPH 1994 Proceeding of the 21st Annual Conference on Computer Graphics and Interactive Techniques (1994)
16. Osgood, C.E., Suci, G.J., Tannenbaum, P.H.: The measurement of meaning. University of Illinois Press, Urbana (1957)
17. Mori, M.: Bukimi no tami, The uncanny valley. Energy 7, 33–35 (1970)
18. Ham, J., Midden, C.: A Persuasive Robotic Agent to Save Energy: The Influence of Social Feedback, Feedback Valence and Task Similarity on Energy Conservation Behavior. In: Ge, S.S., Li, H., Cabibihan, J.-J., Tan, Y.K. (eds.) ICSR 2010. LNCS, vol. 6414, pp. 335–344. Springer, Heidelberg (2010)
19. Siegel, M., Breazeal, C., Norton, M.I.: Persuasive robotics: The influence of robot gender on human behavior. In: IROS 2009, pp. 2563–2568 (2009)
20. Fogg, B.J., Nass, C.I.: Silicon sycophants: the effects of computers that flatter. International Journal of Human Computer Studies 46, 551–561 (1997)
21. IJsselsteijn, W.A., de Kort, Y.A.W., Midden, C., Eggen, B., van den Hoven, E.: Persuasive Technology for Human Well-being: Setting the Scene. In: IJsselsteijn, W.A., de Kort, Y.A.W., Midden, C., Eggen, B., van den Hoven, E. (eds.) PERSUASIVE 2006. LNCS, vol. 3962, pp. 1–5. Springer, Heidelberg (2006)
22. Shinozawa, K., Naya, F., Kogure, K., Yamato, J.: Effect of Robot's Tracking Users on Human Decision Making. In: Proceedings of 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems, Sendai, Japan, pp. 1908–1913 (2004)
23. Iio, T., Shiomi, M., Shinozawa, K., Akimoto, T., Shimohara, K., Hagita, N.: Entrainment of Pointing Gestures by Robot Motion. In: Ge, S.S., Li, H., Cabibihan, J.-J., Tan, Y.K. (eds.) ICSR 2010. LNCS, vol. 6414, pp. 372–381. Springer, Heidelberg (2010)

BEHAVE: A Set of Measures to Assess Users' Attitudinal and Non-verbal Behavioral Responses to a Robot's Social Behaviors

Michiel Joosse[1], Aziez Sardar[1], and Vanessa Evers[2]

[1] Informatics Institute, Faculty of Science, University of Amsterdam, PO Box 94323, 1090 GH Amsterdam, The Netherlands
{mjoosse,aziez}@science.uva.nl
[2] Department of Electrical Engineering, Mathematics and Computer Science, University of Twente.
PO Box 217, NL-7500 AE Enschede, The Netherlands
v.evers@utwente.nl

Abstract. Increasingly, people will be exposed to social robots. In order to inform the design of behaviors for robots that share domestic and public spaces with humans, it is important to know what robot behavior is considered as 'normal' by human users. The work reported in this paper stems from the premise that what would be perceived as socially normative behavior for robots may differ from what is considered socially normative for humans. This paper details the development of a set of measures, BEHAVE, for assessing user responses to a robot's behavior using both attitudinal and physical responses. To test the validity and reliability of the BEHAVE set of measures, a human robot interaction experiment was conducted in which a robot invaded the personal space of a participant. Based on the results from this evaluation, a final set of BEHAVE measures was developed.

Keywords: Human-robot interaction, proxemics, humanlike robots, avoiding behaviors.

1 Introduction

In the future, we can expect that more and more people will encounter robots and may need to interact or work with robots on a regular basis. Where robots were previously confined to industrial environments such as car manufacturing plants, future robots are envisioned to operate in the home [15] and in public spaces. The latter are very different environments for human robot interaction [12]. Industrial environments such as a car assembly line are designed to optimize the robots' operations. In contrast, robots in domestic environments will need to operate in a world specifically designed for humans. As people start sharing their social spaces with robots, the question arises about how they should interact socially. Human social interaction is governed by social norms [12] that dictate what distance to keep, when to acknowledge someone's

B. Mutlu et al. (Eds.): ICSR 2011, LNAI 7072, pp. 84–94, 2011.

presence, engage in eye-contact, approach, start speaking, smile and so on. Social norms differ across cultures and even for people of the same cultural background, interpretation of social behaviors is often fraught with misunderstanding [16].

It is yet unclear whether human social norms are transferable to human robot interaction. Even though there is some initial research that investigates the effect of social normative behavior displayed by robots on human responses to the robots. For instance, [18] investigated US and Chinese responses to explicit and implicit communication, such research has not yielded a complete set of socially normative behaviors for robots that interact with humans.

We aim to contribute to the knowledge in this area by identifying a comprehensive overview of socially normative behaviors for robots. The work reported in this paper concerns the development of a set of measures to evaluate human responses to robot behaviors in order to assess whether they are socially normative. Our aim is to be able to measure people's subjective attitudinal as well as more objective behavioral responses to robot behaviors. With these measures we will then identify which behaviors are most appropriate for human robot interaction. The remainder of this paper will first explain how we developed an initial set of measures. We will then detail how we evaluated the validity and reliability of the measures in an experiment investigating the socially normative behavior of personal space invasion. Finally, we will provide the final set of measures that can be used to measure subject responses to robot social behavior to determine which behaviors are socially normative for human robot interaction.

2 Theoretical Background

2.1 Relevant Measures to Study Human Responses to Social Behaviors of Robots

When evaluating whether a robot adheres to human social norms we have chosen the following approach:

1. From social psychological and behavioral science literature we identify human social behavioral norms (for instance preferred interpersonal space).
2. In an experimental setting, we expose subjects to a human confederate as well as a robot displaying this behavior and either adhering to or violating a specific social norm.
3. Measuring the humans' subjective (attitudinal) and objective (behavioral) response to determine whether adhering to or violating the social norm influences subjects' attitudes and behaviors positively or negatively.

2.2 Attitudinal Response

In order to measure the attitudinal and behavioral responses, we derived a set of measures based on literature which we will discuss here. Attitudinal measures include perceived human likeness of the robot, attitudes toward robots, trust in the robot, perceived social skills of the robot, and physical and social attraction of the robot. Behavioral measures include body responses such as leaning away, stepping away and facial feature responses such as smiling or looking scared. The origin of each measure is detailed below.

Previous studies have found that humanlike robots elicited different responses than more mechanical looking robots because people saw them as more human or because people had higher expectations of more human-looking robots [15]. It is possible that people expect higher conformity to social norms from more humanlike robots due to the closer human resemblance. One of the measurements to include is therefore the **perceived human-likeness** of the robot. For our set of measures we included a scale devised by Ho & MacDorman [8]. This is a 7-point Likert-scale consisting of six items, for example 'human-made – humanlike'. The only adjustment we made was re-adding the item deleted by Ho and MacDorman "genderless – male or female".

When evaluating people's responses to robots, it is important to consider whether people are predisposed to like or dislike robots in general. Results of a study by Wang et. al. [19] suggested that a more negative attitude toward robots may influence a person's tendency to adhere to a robot's recommendations. In order to measure whether people have a generally negative or positive attitude to robots we included the Negative **Attitude toward Robots** Scale on a 7-point Likert-scale as reported by Nomura. [13]. We have excluded three out of the fifteen original items, these items seamed redundant and somewhat ambiguous. Thus leaving the twelve items as reported in the results section.

Trust is an important factor in the use of social autonomous systems [3]. Because trust conveys a lot about the users' attitudinal response towards the robot, we included a measure of trust: the 7-point Likert- Source Credibility Scale [9] developed by McCroskey, et.al. The 3-item subscale 'competence' was not used. These questions concerned the fulfillment of a task, which is not always relevant for interaction with service robots in domestic/public environments. Because of the same reason the two 'communication items' in the subscale 'extroversion' were not used either. The subscales 'sociability', 'composure' and 'character', with items like "calm-anxious" and "tense-relaxed" were adapted to measure responses to a robot.

Because **likeability and attraction** are found to greatly influence the outcome of human as well as human-robot interaction [11], [14] we included items from the Interpersonal Attraction scale, developed by McCroskey & McCain [10] to measure the likeability of the confederate. This originally is a 15-item 7-point Likert-scale. We used the original 10 items from the 'social-' and 'physical attraction' subscales and adapted the questions to measure responses to a robot, for example: "I think the robot could be a friend of mine".

Perceived social skills are also considered important, because higher social skills could lead to higher conformity to social norms. To measure these skills, we included a five-item scale developed by Wish & Kaplan [20] to measure social skills. This 9-point bipolar scale was modified to a 7-point Likert scale.

2.3 Non-verbal Behavioral Response

In order to measure non-verbal behavioral responses more objective by analyzing video material of people interacting with robots, we developed a tool with which coders are able to evaluate the videos. This tool consists of items concerning: immediacy cues, facial expressions, body movement and overall behavior of the user.

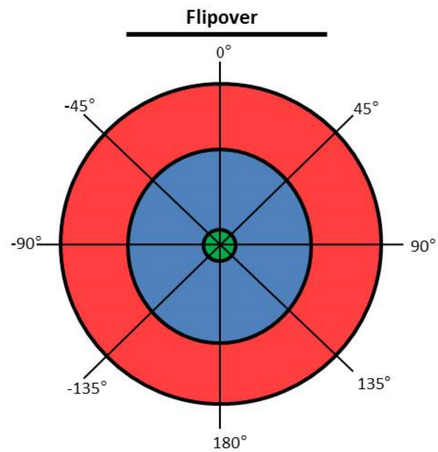


Fig. 1. Step distance and step direction

The immediacy cues consist of step distance and step direction. We developed a "wheel", in which the step direction could be given in eight different degrees as depicted in figure 1. The distance could be either no step, a step within one's intimate zone (less than 45 cm) or a step outside the intimate zone.

Black & Yacoob [2] defined motion cues for each of the six universal emotions happiness, sadness, surprise, fear, disgust and anger. We included the items as 5-point Likert-scaled questions in order to check if these emotions appeared, for example "raising mouth corners". We hypothesized that gaze, staring eye contact would be important, as it has several communication functions, like establishment and recognition of a social relationship [1]. This includes smiling, by raising one's mouth corners. Therefore, we included eight items, for example: "participant looked away", "participant made eye-contact" and "mouth corners raised".

Body posture is also indicates someone's interest. Guerrero defined leaning and touching as immediacy / involvement cues [5]. We included three items, based on the work of Guerrero, these were "the participant leaned away from the confederate/ robot", "participant leaned towards / from the confederate/robot" and "participant was touching him/herself".

The overall emotions were coded by the items defined by Guerrero [5]. We included five 5-point Likert-type items, with which we used the coders' ability of determining the mental state of mind of the user. Examples of these items are: "anxious – calm", "distracted – focused" and "restless – still".

3 An HRI Experiment to Test the Validity and Reliability of the Measures

In order to test the validity and reliability of our measures we administered them in a human-robot interaction experiment. The experiment concerned the socially normative behavior of maintaining personal space, an specific example of social normative behavior is interpersonal distance. People keep what they see as an appropriate distance from each other. This leads them to stand further apart when standing in a big, open room than in a crowded elevator. This distancing behavior is called proxemics, first defined by Edward T. Hall as "the physical and psychological distancing from others"[6]. The personal space cannot be invaded without causing some sense of discomfort [6, 7]. Four different personal space zones (See table 1) have been defined which tend to hold true for most people, although there are factors which influence proxemics, such as gender and social (cultural) norms [1].

Table 1. Personal space zones as defined by Hall. [6]

Personal space zone	Range
Intimate zone	0 - 0.45 m.
Personal zone	0.45 - 1.2 m.
Social zone	1.2 - 3.6m.
Public zone	3.6m +

Recently, Mumm & Multu [11] carried out a study where subjects were instructed to retrieve something from behind the robot that was positioned in the centre of a room. The robot either followed the subject with its gaze (turned its head to follow the movement of the subject as they walked past the robot to retrieve the item at the back of the robot) or kept its gaze fixed to a specific point in the room. They found that subjects kept a larger distance from the robot in the condition where the robot followed the subject with its gaze. The findings suggest that a robot that shows more humanlike behaviors or that seems aware of the humans around it is given more personal space. Even though this is an extrapolation from their conclusions, it seemed as if the subjects in Mumm and Mutlu's study were adhering to a social norm to give more room to a being that was aware of their presence. Other research found that women maintained a larger distance from robots than men when the robot was looking at their face [17]. Similarly, a robot's appearance was found to influence proxemic behaviors. In research by Syrdal et.al. [15] people allowed a mechanical robot to approach them closer (57 cm.) than a humanoid robot (52 cm.) with more anthropomorphic attributes. This could be due to the fact that humanlike robots lead to higher expectations of conformity to social norms. No difference between men and women was reported.

Another factor that could influence the (dis)comfort of the personal space invasion is approach speed. In earlier work from Dautenhahn & Walters a speed of 0.4 m/s was considered a good speed, if perhaps a little too slow [4] when approaching a seated person. On the contrary, a speed of 1.0 m/s was considered as being too fast [18], when approaching a standing person.

For the experiment where the BEHAVE measures were administered, we expected that people would have higher expectations of humans than robots concerning adhering to social norms. We therefore thought that when a human invaded a subject's personal space compared with when a robot would invade the personal space, people would have more negative attitudes toward the human and would display more avoiding behaviors (negative body and facial behaviors). We also expected this effect would be stronger in the case that the approach was faster than what is considered a socially normative approach speed.

3.1 Participants

A total of 92 participants (52 men and 40 women) participated in this study, aged between 18 and 70. (M=24.8, SD = 9.5). 85% of the participants were from the Netherlands, 7% from other European countries and 8% from countries outside Europe. 79% of the participants indicated they (previously) owned a pet. Of the 47 participants who interacted with the robot, 42% of the participants indicated that they had no prior experience with robots and 12% had built robots themselves.

3.2 Methods

We conducted a controlled 2 (human confederate vs. robot confederate) x 2 (comfortable approach vs. uncomfortable approach) between-group laboratory experiment. Depending on the condition, either a robot or a human approached a subject who was viewing a poster and subsequently, the agent invaded the personal space of the subject. The subjects' behaviors were videotaped and later analyzed. The subjects' attitudes were assessed by a post session questionnaire. Only the participants who interacted with the robot were given the human likeness [8] and attitude towards robots questions [16].

Fig. 2. Modified Nomad Scout

The robot used was a modified Nomad robot (see figure 2), with a height of 140 cm. and a diameter of 40 cm. The robot was operated with a joystick from an adjacent room. Depending on the gender of the participant, the confederate was either a men or woman, of average size and posture and wore a white shirt. The approach speed was either comfortable, normal (0.4 m/s), or uncomfortable, fast (1.0 m/s). Behavioral responses were collected using two video cameras: one positioned above the flipover with poster and one in the corner of the experiment room. See figure 3 for the experiment lab layout.

3.3 Experimental Procedure

After being welcomed, informed of the procedure and having signed a consent form, the participants were led to the experiment room. In the corridor they were told that they should not mind the technical equipment (robot condition), or the "other participant" (human condition). Invasion of personal space was not mentioned to the participants. During the experiment, the participants were instructed to search for figures on a poster. After a minute, the human or robot confederate would move towards the participant, and invade the participant's personal space zone. After entering the intimate personal space zone (45 cm.), the agent would turn its "head" towards the participant's poster. The behavior of the participant was observed, after which a debriefing consisting of a questionnaire and brief interview questions took place.

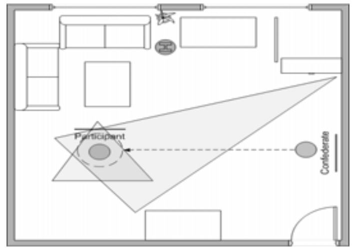


Fig. 3. Experiment lab layout

The video data from our two cameras (see figure 3), the behavioral measurements, were coded by the categories facial expression, intimacy cues and immediacy. Each pair of videos was coded by three coders. There were a total of 30 different coders. None of the coders knew the participants, the coding was done anonymously. The Intraclass Correlation Coëfficient (ICC) was used to calculate intercoder reliability.

4 Results

We will report the reliability of the measures and the final scales, results of the experiment itself will be published elsewhere. Together, the final scales as reported in table 2 and table 3 were included in the final BEHAVE measurement tool.

Table 2. Attitudinal measures: final items and their reliability

Attitude towards robots - NARS [13]	Chronbach's $\alpha = .708$
Subscale Interaction:	
I would feel uneasy if I was given a job where I had to use robots.	
I would hate the idea that robots or artificial intelligences were making judgments about things.	
I would feel very nervous just standing in front of a robot.	
Subscale Social:	
I would feel uneasy if robots really had emotions.	
Something bad might happen if robots developed into living beings.	
I feel that if I depend on robots too much, something bad might happen.	
I am concerned that robots would be a bad influence on children.	
I feel that in the future society will be dominated by robots.	
I feel that in the future, robots will be commonplace in society.	
Subscale Emotion:	
I would feel relaxed talking with robots.	
If robots had emotions, I would be able to make friends with them.	
I feel comfortable being with robots.	
Attractiveness of the robot- Interpersonal Attraction Scale (physical) [10]	Chronbach's $\alpha = .820$
I think the robot is quite handsome	
The robot is very sexy looking	
I find the robot very attractive physically	
I don't like the way the robot looks	
The robot is somewhat ugly	

Table 2. (*Continued*)

Likeability of the robot - Interpersonal Attraction Scale (social) [10]	Chronbach's
I think the robot could be a friend of mine	α = .777
It would be difficult to meet and talk with the robot	
The robot just wouldn't fit into my circle of friends	
The robot and me could never establish a personal friendship with each other	
I would like to have a friendly chat with the robot	
Human likeness - Ho & MacDorman [8]	Chronbach's
Artificial – Natural	α = .821
Human-made – Humanlike	
Without definite Lifespan – Mortal	
Inanimate – Living	
Mechanical Movement – Biological movement	
Synthetic – Real	
Genderless – Male or Female	
Trust in the robot- Social Credibility Scale [9]	Chronbach's
Good natured - irritable	α = .789
Cheerful-gloomy	
Poised-Nervous	
Tense-Relaxed	
Calm-Anxious	
Dishonest-Honest	
Unsympathetic-Sympathetic	
Good-Bad	

The internal reliability of the attitudinal measures was high, except for one, the social credibility scale (Chronbach's α = .48). After removing the items "friendly-unfriendly" and "timid-bold", the reliability became acceptable (Chronbach's α = .789). The final items and their internal reliability are included in table 3. The internal reliability of the social skills measure was very low (α < 0.5); therefore this measure was excluded from BEHAVE.

For the behavioral variables to measure avoidance behavior, included in table 3, intercoder reliability was low for many of the items. The items with a high α (indicating that the coders may have interpreted the responses roughly the same) were combined into one independent variable, called "avoiding behavior" (Chronbach's α = 0.836); a variable from 1 to 5, with 1 being "no avoiding behavior" and 5 "a lot of avoiding behavior". The ICC of the step direction variable was high (.829, α = .936), as well as that of the step distance (.878, α = .956).

Table 3. Avoidance behavior measure: final items and reliability

Item	
Participant made eye-contact	α = .836
Did the participant say anything the moment that PSI occurred?	
Participant laughed out loud	
Participant was distracted from his/her task	
Mouth corners raised	
Participant leaned away from the confederate/robot	
Eyebrows raised	
Eyes open wide (to expose more white)	

5 Discussion and Conclusion

In this paper, we presented a tool for measuring socially normative robot behavior. The variables human likeness, attraction, likeability and attitude towards robots are all included to be able to explain differences in response to a robot's behavior in an experimental setting. The variables trust and avoidance/engaging behavior are the dependent variables of interest in order to measure whether certain robot behaviors are considered socially normative. We believe the set of measures as well as the video-coding method will prove to be an extremely valuable resource in the near future as it allows researchers to assess the quality of their robot's behavior in an experimental setting. This indicates, in our opinion the extent to which the behavior is experienced as socially normative by the users.

The scales showed high internal reliability and the question that remains is whether the scales are valid and indeed measure perception of socially normative behavior. Because the trust in robots scale is based on the previously validated McCroskey source credibility scale, its validity is more easily argued for. The behavior scale is more difficult to assess the validity of. When looking at the final scales, the questions seem to address (positive) engagement or (negative) avoidance behavior. Even though this is the first study to assess the BEHAVE set of measures, we think that engagement is a valid indication of users' behavioral responses to robots and we expect that future research will confirm this. Even though we believe these results are sound, replication of this experiment would prove useful to determine the external reliability.

We adopted both attitudinal and behavioral measures. Of the behavioral measures the immediacy cues evaluated by the coders had the highest intercoder reliability. We believe that it is very difficult for coders to assess subtle facial feature movements and possibly technological advances in facial recognition can be used in the future to objectively detect emotional responses.

The attitudinal measures (NARS, attraction, likeability, human likeness and trust) were very reliable. We tried to measure the social skills but it resulted in an incorrigible low Alpha. The cause of this was probably because the questions were multi-interpretable. In spite of this, we still think it is imperative to measure the believed social skills of the social robot. Future work will focus on developing a set of items to measure the perceived social skills of the robot as it will be an indication of how socially able users found the robot and possibly, how socially normative the robot is.

References

1. Argyle, M., Dean, J.: Eye-Contact, Distance and Affiliation. Sociometry 28(3), 289–304 (1965)
2. Black, M.J., Yacoob, Y.: Tracking and Recognizing Rigid and Non-Rigid Facial Motions using Local Parametric Models of Image Motions. In: Fifth International Conference on Computer Vision, pp. 374–381 (1995)
3. Cramer, H., Evers, V., van Someren, M., Ramlal, S., Rutledge, L., Stash, N., Aroyo, L., Wielinga, B.: The effects of transparency on trust and acceptance in interaction with a content-based art recommender. User Modeling and User-Adapted Interaction 5(18), 455–496 (2008)
4. Dautenhahn, K., Walters, M., Woods, S., Koay, K.L., Nehaniv, C.L., Sisbot, E.A., Alami, R., Siméon, T.: How I Serve You? A Robot Companion Approaching a Seated Person in a Helping Context. In: Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction, pp. 172–179. ACM Press, New York (2006)
5. Guerrero, L.K.: Observer Ratings of Nonverbal Involvement and Immediacy. In: Manusov, V.L. (ed.) The Sourcebook of Nonverbal Measures. Psychology Press, Philadelphia (2004)
6. Hall, E.T.: The Hidden Dimension. Doubleday, New York (1966)
7. Hayduck, L.: Personal Space: Where We Stand Now. Psychological Bulletin 94(2), 293–335 (1983)
8. Ho, C.C., MacDorman, K.F.: Revisiting the uncanny valley theory: Developing and validating an alternative to the Godspeed indices. Computers in Human Behavior 26(6), 1508–1518 (2010)
9. McCroskey, J.C., Jensen, T., Valencia, C.: Measurement of the credibility of peers and spouses. Papers presented to the International Communication Association, Montreal (1973)
10. McCroskey, J.C., McCain, T.A.: The Measurement of Interpersonal Attraction. Speech Monographs 41, 261–266 (1974)
11. Mumm, J., Mutlu, B.: Human-Robot Proxemics: Physical and Psychological Distancing in Human-Robot Interaction. In: Proceedings of the 6th International Conference on Human-robot Interaction, pp. 331–338 (2011)
12. Nakauchi, Y., Simmons, R.: A Social Robot that Stands in Line. Autonomous Robots 12, 313–324 (2002)
13. Nomura, T.: Measurement of Negative Attitudes Towards Robots. Interaction Studies 7(3), 437–454 (2006)
14. Patterson, M.L.: An arousal model of interpersonal intimacy. Psychology Review 83(3), 235–245 (1976)
15. Syrdal, D., Dautenhahn, K., Walters, M.L., Koay, K.L.: Sharing Spaces with Robots in a Home Scenario – Anthropomorphic Attributions and their Effect on Proxemic Expectations and Evaluations in a Live HRI Trial. In: Proceedings of the 2008 AAAI Fall Symposium, pp. 116–123 (2008)

16. Taylor, O.L.: Cross Cultural Communication: An Essential Dimension of Effective Education. The Mid-Atlantic Equity Center, Maryland (1987)
17. Takayama, L., Pantofaru, C.: Influences on proxemic behaviors in human-robot interaction. In: Proceedings of the International Conference on Intelligent Robotic Systems: IROS 2009, St. Louis, MO, pp. 5495–5502 (2009)
18. Wang, L., Rau, P., Evers, V., Robinson, B., Hinds, P.: Responsiveness to Robots: Effects of Ingroup Orientation & Communication Style on HRI in China. In: Proceedings of Human Robot interaction Conference HRI 2009, San Diego, March 11-13, pp. 247–248. ACM Press, New York (2009)
19. Wang, L., Rau, P., Evers, V., Robinson, B., Hinds, P.: When in Rome: the role of culture & context in adherence to robot recommendations. In: Proc. of the 5th ACM/IEEE International Conference on Human-Robot Interaction, pp. 359–366. ACM Press, New York (2010)
20. Wish, M., Kaplan, S.T.: Toward an Implicit Theory of Interpersonal Communication. Sociometry 40(3), 234–246 (1977)

Initial Formation of Trust: Designing an Interaction with Geminoid-DK to Promote a Positive Attitude for Cooperation

Elizabeth G. Dougherty and Henrik Scharfe

Department of Communication - Aalborg University,
Kroghstraede 3, Aalborg East
{edough10,scharfe}@hum.aau.dk

Abstract. In this paper we present a study of how touch may be used as a way of inducing trust in the meeting with a teleoperated android, the Geminoid-DK. The use of haptics with Gestalt-based 'Balance Theory' is used as a persuasive design of the interaction between human and robot. Balance theory includes both cognition and affect as factors influencing motivated behavior, subsequently effecting judgments of trustworthiness. Haptic interactions have both conscious and unconscious implications for the receiver in this interaction, which we demonstrate by an experimental set-up, as well as through questionnaires and interview.

Keywords: Geminoid, android, affect, Balance Theory, haptic communication, persuasion, teleoperation.

1 Introduction

We adapt the Balance Theory model to explain how beliefs and attitudes toward a geminoid as a communication medium can determine levels of trusting to develop successful business exchanges. The contributions are two-fold, we propose a person's trust in a geminoid greeting is based on the balance model developed by Fritz Heider [8]. We are looking at it from a business-like situation, but also from a more general human approach within the robotic medium. We propose that people, upon meeting the geminoid, may form an attitude which can be positively influenced. Likewise, the formation of this overall opinion towards the geminoid should be made in an assisted environment to support the positive acceptance of this form of mediated technology.

To fulfill the mentioned contributions, we discuss why Balance Theory is appropriate. We hypothesize that designing the situation in a certain way will create more trust with the robot. Therefore, we focus on analyzing certain factors that influence behavior. We present the attitudes toward the geminoid and discuss the concept of trust with a focus on human-robot interaction. We introduce the theoretical foundations in sections 2 and 3 and in section 4 we apply this to communication with the geminoid robot. We present the empirical study in section 6, including concluding remarks.

B. Mutlu et al. (Eds.): ICSR 2011, LNAI 7072, pp. 95–103, 2011.

2 Introduction to Balance Theory

Balance theory was developed in the 60's by Fritz Heider and has been studied extensively since. This theory defines the phenomenon of people within a group of three or more actively creating a balance when the relationships are mismatched. The mismatching will eventually lead to one of the persons to accumulate and change. Here, we are looking at Balance Theory and if it can be applied to non-verbal communication to communicate attitudes, emotions, and haptic experiences. This theory is applied to both the assistant's behavior toward the geminoid overall, and the visitor's behavior as part of a new situation of conducting business with a robot. In this study, we consider the touch that is normally part of care-giving assistance, including touches to the arm and hand area, as well as handshaking. Judgment of the relationship of the assistant to both the robot and the new person will, in time, be 'balanced' and the triadic relationship will be consistently positive (or negative). The use of touch supports the positive emotions and interaction in the triad. With the warm human touch of the assistant to reassure the new person, their assessment of touching the robot, with this theory, should be balanced over time. In [14], we find an example of Balance Theory applied to communication involving humans and an android, with special emphasis on gaze. In this study, we apply the theory with a focus on touch as a communicative factor.

3 Touch, Persuasion, and Attitudes

We follow the tradition of persuasive technology where the aim is to change beliefs, attitudes and behaviors in people with the use of technology. As described by B.J. Fogg [6], persuasion requires intentionality, which is the planned persuasive effect within an interaction. Designing the interaction with a geminoid tries to direct behavior by having an emotional involvement with the receiver being more open to the situation as a result. We are looking for a heightened amount of presence while in the mediated interaction with the geminoid. Presence is generated from "sensory input, mental processes, and past experiences." O'Keefe (p. 15). In human computer interaction, people have the tendency to anthropomorphize computers. With social cues like the life-sized facially realistic Geminoid DK, the effect is intensified as a social actor. A social actor, according to B.J. Fogg [6], is an interactive technology that is responded to socially, as though it were a living being. Touch is persuasive in that it is an often overlooked channel of communication. It comes in many forms and can be more powerful than language, especially in its negative form (Burgoon et al., 1996). It can range from being comforting to being anxiety-provoking. An effect of compliance is also made with touch as in the persuasive "Midas Touch" effect [5]. When patrons of a restaurant were touched lightly by the waitress, the patrons gave a significantly higher tip than the no-touch control group. This compliance may be a form of cooperation and a way to show that the waitress bonded with the patron. The assistant in this experiment could be considered part of the hospitality industry like the above waitress, hoping that the people they are greeting and assisting will be comfortable. Like a waitress receives money for their work, here, a trust game is played out in our experiment. These effects have been argued by Rose (1990) that the recipient often assumes that the toucher genuinely likes the participant and trusts him or her which is why the Midas Touch effect occurs. The perception of a positive feeling tends to increase compliance rates. In this explanation by Rose, trust is the base emotion, which increases compliance. Reite (1990) draws a line between touch and stress reduction as the key to the positive response, as touch is used as a stress reducer in early childhood by parents. Both a trust response and a calming effect are important when interacting with a novel social robot.

The drive for intimacy and touch has biological roots meaning that it is something that nearly all humans share. Touching is human's first form of communication. There is a relationship between attachment and intimacy. With bonding there is trust, and this is the aim of many businesses today in an area where technology is both bringing people closer, yet farther apart because it is still in the process of developing. Teleoperated androids have appealing convenience qualities for business, but if the people interacting with them are not able to bond and form trusting relationships through androids, then the applications for this technology is limited. Affectional bond is the attraction one individual has for another, John Bowlby [1] states. The psychologist worked on central features of affectional bonding, and he describes the relations as that of a child who has a bond with his or her mother. We regard the assistant position as being analogous to an affectional parent, one who is looked at as the 'safety' person in the situation. This nature of bonding and attachment implicates why balance theory is important. People have a social goal of having others like them approve of them and get affection from others.

4 Touch, Trust and the Geminoid

Touch as a way of building trust in human robot interaction is sometimes mentioned as a prerequisite for successful social interaction between humans and robots [15], has been studied under many conditions. Studies such as reported in [16] show that touch by zoomorphic robots can be effective in certain circumstances. In this study, we use an android, and focus on the effect of letting test subjects touch the robot, prior to participating in a joint task with it. Teleoperated robots in the workplace allow for business traveling and working at home while maintaining a presence at the office. We were interested in seeing how important the experience of actually touching the robot was, and if it was possible to design the interaction with the robot such that trust could be build efficiently. The physicality of the robot opens up the opportunity of utilizing haptics as a supportive dimension of communication. And by designing a physical interaction that involves touch, we focus on a main characteristic of the geminoid – the physical presence – as opposed to non-robotic telepresence systems. For the study we use the Geminoid DK. As the other android in the geminoid series, the DK has a very high degree of physical likeness with a living person, and it is designed to operated at a distance.

In the classic study on touch by Fisher, Rytting and Heslin on tactile stimulation of a fleeting touch to the hand, library clerks gave a momentary touch to people checking out their books. The touch occurred while passing the library card back to the subject. This touch was not consciously noticed by many of the people yet there was a significant effect of the touch. The findings pointed to positive affect when the mild tactile stimulus was given. Further, the affective state was thought to determine the evaluation of associated stimuli as described by Byrne-Clore as positive emotions being generalized within an environment (1970). When the subjects filled out a questionnaire, the clerk was rated more favorably and a more positive affect was experienced by those in the touch condition. This casual touch had positive effect on all touched individuals. Significantly, the recipients who were not consciously aware of being touched had similar measures to those who were conscious of being touched. Generally, the touch created a positive effect whether it was consciously received or not. Byrne and Clore (1970) furthered their arguments on generalization of positive emotions in that touch mediates evaluation judgments to points of association within the setting. We emphasize that a stronger effect of positive emotions caused by touch impacts the entire situation.

'The Initial Introduction' was created by Kahn et al. for the autonomous robot Robovie because they “know ...that when we meet another person for the first time that there is often some initial awkwardness, or at least a lack of knowledge of one another.” They acknowledged that all cultures have rituals for dealing with initial moments in an interaction. “We asked ourselves: Why not expect that something similar could emerge when people meet social robots for the first time?” Their protocol started with the experimenter saying to the participant, “I have someone I would like you to meet”. Then the robot would initiate the conversation with the participant, saying hello and asking if he or she would like to shake its hand. Then it would ask how the participant is doing and end with closing remarks (pg. 100). Having the robot initiate the “hello” sequence with the participant and stretching out its hand to the participant is a persuasive way to establish that the participant can learn something from the robot. The robot is in the role of the care-giver and the participant is the receiver. This makes the participant more open to bonding with the robot, and gives a sense of security to the participant if they have not interacted with a robot before.

We would like to put forth the following hypothesis: In a human-robot situation, an assistant who touches a subject in a structured greeting protocol will create positive emotions through the tactile stimuli.

Handshaking is a common practice in business which includes social cues and feedback meanings. It is the time when people make their first impressions and expectations about one another. It is part of a western approach to interacting with the geminoid in a business situation, where handshaking is part of the initial stage of forming a business relationship. Nonverbal behaviors may have multiple meanings, but a handshake is a simple way for two people to be within a close proximity and to have a direct eye gaze and touch, all behaviors that encourage positive emotions and approach-behavior. This exchange can clarify the status of the interpersonal relationship. According to Burgoon, “Forward lean and touch conveyed greater intimacy/attraction/trust and greater involvement than backward lean and absence of touch.” Similarly, that “greater nonverbal immediacy, closer proximity, more gaze, forward body lean, vocal and gestural animation, faster (tempo) conveyed not only greater involvement and affection but also greater receptivity/trust.” (p. 237). All of the above mentioned behaviors and effects are part of a handshake greeting.

Lack of knowledge of teleoperated robots is problematic as people are not certain whether the robot is autonomous, or if a human being is speaking through the robot. The lack of experience and not being a typical part of our current culture, affected the participants view of the purpose of interacting with the robot. If there is not an agreement that the robot would serve a further purpose for the participant, then he or she is less apt to try to get comfortable with it. This is effecting the trust intention, meaning the intention to engage in trusting behaviors.

5 Empirical Study

In this section, we investigated how the behavior of an assistant affects a participant's impressions toward an android (a third person). The second person used a touch or no touch condition with a verbal greeting as a communicative behavior towards the subjects.

5.1 Subjects

A total of 12 adults or students enrolled at a major university in the northern part of Denmark participated in the experiment. Half of the participants (n=6) were randomly assigned to one of the two conditions. The first condition was performed with six subjects (three males and three females.) The second condition was also performed with six subjects (three males and three females.) All subjects signed a consent form.

5.2 Apparatus

A life-sized tele-presence android called Geminoid DK was used in the experiment. Its realistic physical characteristics are an exact copy of its original operator. Movements of brows, eyes, cheeks (indicating smiling), and shoulders (indicating breathing) are preprogrammed, but can be overruled by the operator at run-time. Movement of head and torso is controlled by an operator also at run-time, and lip movement of the android is synchronized to the speech of the operator. Although the setup requires many technical choices, the actual interface of operating the geminoid is fairly straight-forward, which allows the operator to maintain a strong focus on the interaction the test room. As the movements are restricted to the face and upper torso of the android, it is unable to initiate a proxemic experience or tactile interaction with the test subjects. Audio and video of all interlocutors were digitally recorded.

5.3 Environment

The experiment was conducted in a typical single-sized office room with a window and a door to the hallway. The android was seated behind a desk, dressed in business clothing wearing a dark colored shirt and tie, suit pants and shoes. Upon the desk was a computer, books and a business newspaper; the laboratory had a business-like décor. There were some visible wires behind the android, and three visible video cameras on tripods.

Speakers were placed behind the android to relay the speech of the operator. A microphone was placed in the middle of the desk. No attempt was made to conceal wires and other technical equipment, but all these necessities were arranged as neatly and non-obtrusively as possible. The assistant sat at the side of the rectangular desk and the participant sat across from the android in an office chair.

Fig. 1. Left: The participant faces the geminoid, the assistant in the middle. Right: The operator in the control room. Next to him, the researcher can monitor audio and video while preparing for the exit interviews.

5.4 Procedure

An assistant acts in both of the conditions: one with tactile touch and the other without tactile touch. First, the participant was told the rules of the trust game in a room separate from the laboratory with the android, and asked if he or she had any questions. Then the participant went into the laboratory to see the geminoid for the first time for 30 seconds while the geminoid only had the autonomous movements activated. The subject was led out of the laboratory and immediately greeted by the assistant and asked to reenter the laboratory again with the assistant. The subject exchanged introductions with Geminoid DK with a verbal greeting, with or without the tactile greeting. After the greeting, the assistant conducted a 'trust game' between the participant and Geminoid DK. The assistant did the physical actions for the Geminoid DK during the game. Real Danish money was used for 10 monetary units (MU). Finally, the participants completed a survey questionnaire and a short interview with the experimenter.

Because the Geminoid's hand has a waxy cold feel to it, and does not move autonomously, the participants were asked to put their hands on the clothed shoulder of the Geminoid. Physical feedback to the participant was in the form of the breathing torso movements. Though touching on the shoulder may be considered vulgar in some cultures, a shoulder-touch is a comparable alternative to a handshake in Western culture.

5.5 Conditions

The participant was introduced to Geminoid DK by the assistant in a touch or a non-touch condition. In both conditions initially, names and a few sentences were exchanged. The greeting stage without touch had the option of verbally continuing between the participant and Geminoid DK for up to 5 minutes. The participant was instructed not to touch Geminoid DK. If the participant attempted to touch the robot, they were instructed not to. In the touch condition, the assistant would touch the participant on the arm guiding him or her to Geminoid DK. While still touching the participant's arm, the assistant instructed the participant to touch Geminoid DK's hand. Then the assistant instructed the participant to sit in the chair opposite Geminoid DK to conduct the trust game. When the trust game was over, the assistant would lead the participant to another room to fill out a questionnaire and have a short interview.

5.6 Measures

5.6.1 Trust Game

Two measures were conducted in this experiment. The 'trust game' is the first measurement, which is a socioeconomic game measuring trust via decision making including risk-taking. There are two roles: one of investor and one of trustee (Geminoid DK) where the investor (the participant) is given a sum of money of 10 MU, roughly equivalent to 100 USD, and told that if they choose to give none, all or some of the MU to the trustee, the assistant will triple the money. The trustee is told that they can send an amount back between zero and the full amount of money. The trustee's final payoff is the initial amount plus the tripled transfer of the investor minus the back transfer to the investor. The participants will not keep the money in this experiment.

5.6.2 Questionnaire

The second measure was two altered questionnaires developed by B. Witmer and M. Singer, Presence Questionnaire (PQ) and Immersive Tendencies Questionnaire (ITQ). These questionnaires were altered to fit in a human-geminoid interaction. Lastly, a general interview about their own explanation of their experiences was notated.

5.6.3 Video Observation

Through recorded video, we were able to observe the interactions of the participant. We looked at mirroring behaviors and positive emotion behaviors such as smiling.

5.7 Results

The results from the questionnaires show that there is little difference between the touch and no touch condition. But the results of the trust game show that there is a significant increase of trust in the touch condition. We suggest that this supports the notion that the subjects were not aware of their increase of trust while rating their trust levels on a questionnaire, but when in a trust situation with the android, they displayed unconscious trusting behavior (during the trust game.)

In the questionnaire, the self-monitored behavior does not differ much between the touch and no-touch conditions. The median answer is 4.7 in the no-touch group and is 5 in the touch group. Moreover, there is only a small difference in the mean amount given to the android in the trust game. The mean in the no-touch group is 4 and the mean in the touch group is 4.16. However, 25% of the subjects in the touch group choose the maximal amount to bet, whereas 16.6% chose the maximal amount in the no-touch group. Thus the touch increases the investors transfer levels in the trust experiment. Therefore the differences between the questionnaire and the trust evaluation are highly significant, suggesting that touch in a human-human-robot interaction affects trust in interpersonal interactions. Touching of the hand reinforces the mechanical aspects of Geminoid DK breaking the suspension of human-like illusion. The enactment of this cultural pattern does not override the corpse-like information that the participants receive. Placement of the participants hand on the shoulder of the robot was more effective than on the hand. The clothing and breathing cycle made the tactile experience more like touching a human. By leaving the hand there for up to a full minute, the participants had close proximity to the robot while it was speaking. This could be interpreted as a threat/confrontation, or a positive way to engage in the robot's presence.

6 Academic Implications

The research reviewed here on touch and trust in an interpersonal interaction shows that there are notable effects on behaviors, judgments and emotions. Specifically, interpersonal touch has both unconscious and conscious effects on trusting responses in a novel situation with a robot. These responses are modulated by compliance and the caring role played by the assistant. Balance theory can be applied to a human-robot interaction with a participant, and a protocol for an optimally designed interaction is a contribution to this new science. Combining human and robot touch can affect people's attitudes and facilitate bonding. The research on why touch is profound in interactions is in its infancy, and has not yet uncovered the reasons behind this phenomenon.

Acknowledgements. Thanks go to Jens Vilhelm Dinesen and Evegenios Vlachos for their help in carrying out the experiment.

References

[1] Bowlby, J.: The Natue of the Child's Tie to his Mother. International Journal of Psycho-Analysis 39, 350–373 (1958)
[2] Burgoon, J.K.: Relational message interpretations of touch, conversational distance, and posture. Journal of Nonverbal Behavior 15(4), 233–259 (1991)
[3] Burgoon, J.K., Buller, D.B., Woodall, W.G.: Nonverbal Communication. McGraw-Hill, New York (1996)

[4] Byrne, D., Clore, C.L.: A reinforcement model of evaluative responses. Personality: An International Journal 1, 103–128 (1970)
[5] Crusco, A.H., Wetzel, C.G.: The Midas Touch The Effects of Interpersonal touch on Restaurant Tipping. International Journal of Hospitality Management 26(4), 1019–1023 (2007)
[6] Fogg, B.J.: Persusaive Technology: Using Computers to Change What We Think and Do. Morgan Kaufmann Publishers, Amsterdam (2003)
[7] Gallace, A., Spence, C.: The science of interpersonal touch: an overview. Neuroscience and Biobehavioral Reviews
[8] Heider, F.: Attitudes and Cognitive Organization. Journal of Psychology 21, 107–112 (1946)
[9] Kahn, P.H., Freier, N.G., Kanda, T., Ishiguro, H., Ruckert, J.H., Severson, R.L., Kane, S.K.: Design Patterns for Sociality in Human-Robot Interaction, HRI 2008, March 12-15 (2008)
[10] McKnight, D.H., Choudhury, V., Kacmar, C.: Developing and Validating trust Measures for E-Commerce: An Integrative Typology. Information Systems Research 13(3), 34–359 (2002)
[11] O'Keefe, J.: Is consciousness the gateway to the hippocampial cognitive map? A speculative essay on the neural basis of the mind. Brain and Mind 15, 59–98 (1986)
[12] Reite, M.: Touch, attachment, and health: is there a relationship? In: Bernard, K.E., Brazelton, T.B. (eds.) Touch: The Foundation of Experience, pp. 195–228. International Universities Press, Madison (1990)
[13] Rose, S.A.: The sense of touch. In: Bernard, K., Brazelton, T.B. (eds.) Touch: The Foundation of Experience, pp. 299–324. International Universities Press, Madison (1990)
[14] Shimada, M., Yoshikawa, Y., Asada, M., Saiwaki, N., Ishiguro, H.: Effects of Observing Eye Contact between a Robot and Another Person
[15] Breazeal, C., Takanishi, A., Kobayashi, T.: Social Robots that Interact with People 2008. Springer Handbook of Robotics, Part G, pp. 1349–1369 (2008)
[16] Evers, V., Winterboer, A., Pavlin, G., Groen, F.: The Evaluation of Empathy, Autonomy and Touch to Inform the Design of an Environmental Monitoring Robot. In: Ge, S.S., Li, H., Cabibihan, J.-J., Tan, Y.K. (eds.) ICSR 2010. LNCS, vol. 6414, pp. 285–294. Springer, Heidelberg (2010)

Minimal Group – Maximal Effect? Evaluation and Anthropomorphization of the Humanoid Robot NAO*

Dieta Kuchenbrandt[1,**], Friederike Eyssel[1], Simon Bobinger[1], and Maria Neufeld[2]

[1] Center of Excellence in Cognitive Interaction Technology (CITEC), Bielefeld, Germany
{dkuchenb,feyssel,sbobinge}@cit-ec.uni-bielefeld.de
[2] Department of Psychology, Dresden University of Technology, Dresden, Germany
maria.neufeld@mailbox.tu-dresden.de

Abstract. How can we increase acceptance and anthropomorphism of robots? In an experiment with $N = 45$ participants, we tested whether categorizing the humanoid robot NAO as an in-group member vs. an out-group member would result in more positive evaluations and higher levels of anthropomorphism of the robot NAO. Results fully support our hypotheses. Moreover, the present findings also indicate that sharing in-group membership with NAO led to greater willingness to interact with robots in general.

Keywords: Anthropomorphism, human-robot-interaction, minimal-group paradigm, robot evaluation.

1 Introduction

Imagine the following scenario: Only several years from now, a robot could be your co-worker on the job. It might also become reality that robot assistants will serve you in elderly care sooner or later. For many people, this vision seems disconcerting, or even unpleasant. What can interdisciplinary research teams from diverse backgrounds such as psychology, design or robotics contribute to make this vision a pleasant one and to increase acceptance of robots in every-day life? In this paper, we present a new approach to facilitate positive reactions toward robots that is based on social psychological research on intergroup relations. More specifically, we investigated whether perceiving a robot as a member of one's own group could increase acceptance and anthropomorphization of the robot.

According to [1], anthropomorphism can be described as "imbuing the imagined or real behaviour of nonhuman agents with humanlike characteristics, intentions, and emotions" (p. 864). Furthermore, in their Three-Factor-Model of Anthropomorphism, [1] suggest that humanizing nonhumans represents an effective strategy to reduce the uncertainty that is often associated with interactions with nonhuman agents. They reason that anthropomorphism serves to satisfy an individual's need to gain mastery and control over one's environment. To illustrate, [2], for instance, have demonstrated that anthropomorphizing technical devices, such as cars or computers leads to reduced stress in human-machine-interaction.

* This research was funded by the German Research Council (COE 277).
** Corresponding author.

B. Mutlu et al. (Eds.): ICSR 2011, LNAI 7072, pp. 104–113, 2011.

To increase anthropomorphism and to facilitate human-robot-interactions (HRI) by reducing stress and uncertainty, one possible approach could be to design robots that appear and behave humanlike [3], [4]. [5], for instance, showed that a robot's ability to express contingent emotional reactions to a human interaction partner significantly contributed to increased anthropomorphic inferences than when the robot merely reacted in a neutral manner. Similarly, the robot's emotional expressivity also ameliorated judgments of the quality of HRI. However, certain problems can be associated with high levels of a robot's humanlikeness: Design-oriented approaches in robotics emphasize the relevance of a robot's physical appearance [6], [7]. A robot's appearance clearly affects the mental models users develop about robots and determines inferences about a robot's functions and capabilities. To illustrate, [3] showed that robots that differed in feminine vs. masculine appearance were gender-stereotyped in terms of warmth and competence and their suitability for gender-stereotypical tasks. Furthermore, [8] proposed that people experience uncanny feelings when confronted with an ambiguous entity that appears neither clearly human-like nor robot-like.

To circumvent problems associated with the design and function of technical devices, we suggest to apply and test social psychological approaches from intergroup research to HRI. Social psychological research has demonstrated that we categorize people according to their membership in key social categories (e.g., age, gender, or ethnicity) [9]. Moreover, depending on whether we perceive others in terms of being in-group or out-group members, intergroup bias might result [10]. That is, people evaluate their own group more positively than an out-group [11], [12]. People even perceive in-group members as more human than out-group members [13], [14], meaning that they dehumanize out-groups. To illustrate, [14] have demonstrated that people ascribe out-groupers less secondary emotions than members of their in-group, whereas no differences in the ascription of primary emotions were found. Secondary emotions (e.g., hope, love, or guilt) have been shown to be applicable only to humans, and thus, represent one core aspect of humanity [14], [15]. In contrast, primary emotions (e.g., happiness, surprise, anger) can be experienced by both humans and animals and thus are not perceived as uniquely human.

Importantly, research has demonstrated that such in-group bias needs not necessarily to be based on meaningful social categories. Instead, in their classic work on the minimal-group paradigm, [9] have shown that people even favor their own group when the groups were randomly formed based on arbitrary features (e.g., the color labels). These effects even emerge despite the fact that participants do not know each other personally and neither do they share a common group history. In short, arbitrary features might serve as cues for subsequent categorization processes and ultimately lead to preferential judgments and treatment of individuals that are perceived in terms of in-group membership.

In line with [16], we propose that people would even treat robots as in-group members, and accordingly, they would show in-group bias toward robots that belonged to an in-group. First evidence for this notion comes from work by [16]. In this study, participants were presented with a robot that ostensibly was a member of their national in-group (a German robot), or they judged the same prototype, but this time, it was said to belong to a national out-group (a Turkish robot). As predicted, participants evaluated the in-group robot more positively and anthropomorphized it more strongly than the out-group robot. Analogous effects were demonstrated for computers and computer agents [17], [18], [see also 19]: For example, when participants worked on a team with a computer (vs. did not do so), participants in the 'team' condition were more likely to cooperate with the computer and evaluated the computer more positively than participants in the control condition [18].

These findings indicate that nonhuman agents indeed can be perceived as members of one's own group and that people tend to show in-group favoritism toward technical systems just like they would toward human counterparts. However, to date, existing work has merely focused on group manipulations that were either personally relevant to the participants' social identity (e.g., ethnicity, [16], [17]) or that included both interdependence between interaction partners and actual human-nonhuman interactions [18], [19]. Accordingly, the effects might have emerged due to pre-existing attitudes toward one's national or ethnic in-group that just transferred to the nonhuman agent or due to the interdependence between the participant and the nonhuman agent, respectively. With the present research, we aimed to address these gaps in the literature. We utilized a minimal-group approach to investigate the effects of merely categorizing a robot as either an in-group or an out-group member on evaluating and anthropomorphizing the robot. In addition, we tested whether perceiving a robot as an in-group member would transfer to a positive bias toward robots in general. Accordingly, we hypothesized that perceiving a robot as an in-group member vs. an out-group member would lead to increased acceptance of the robot, to more anthropomorphic inferences about the robot and, finally, to a greater willingness to interact with robots in general.

2 Method

2.1 Participants

Participants were $N = 45$ German university students (25 male, 18 female, two persons did not indicate gender) with a mean age of 24.81 years ($SD = 5.00$). They were randomly assigned to one of two experimental conditions: Accordingly, they either evaluated a robot that ostensibly belonged to their in-group ($N = 21$) or that belonged to the out-group ($N = 24$).

2.2 Procedure

Participants were tested individually in the laboratory. The experimental procedure served to make participants believe that they would form a group together with the robot NAO (Aldebaran Robotics, [20]) or that they would be part of a group that consisted of other participants of the experiment.

Initially, participant received written instructions about the alleged purpose and the procedure of the experiment. They learned that they would take part in a study on the development of two different language-learning trainings both for robots and humans. Importantly, participants were informed that there would be two groups in the study—each for testing one language-training program—that were labeled with colors: Blue and green. We used colors as group names to make sure that the group categories would not be of any social relevance to the participants. Crucially, participants were told that the robot NAO would belong to the blue group, but not to the green group.

Moreover, participants were informed that they would be randomly allocated to one of the two groups. This manipulation served to create a minimal group situation.

After reading the information sheet, participants were assigned to one of the groups followed by a short interaction phase with NAO [see 21] in which the robot has been introduced to the participants.

Finally, the experimenter asked participants to complete several computerized tasks that were allegedly related to the upcoming language-training test. In fact, however, these tasks contained our dependent measures and no actual language-training test took place. Finally, participants were reimbursed, debriefed and dismissed.

2.3 Experimental Manipulation

In order to manipulate the participants' and the robot's group membership, participants in the *in-group condition* were told that they would belong to the "blue group" together with the robot NAO. Specifically, after the random allocation to one of the two groups, the experimenter said:

You have been assigned to the blue group. This is the group to which the robot NAO belongs. That means several NAO robots have already tested the language training before.

Importantly, the latter information was provided to make sure that participants would not expect having to test the language training in cooperation with NAO. To further enhance the salience of participants' group membership and to assure that they would perceive NAO as an in-group member, the experimenter then continued:

I am now going to present you the robot NAO ***which is a member of your group****.*

In contrast, participants in the *out-group condition* were told that they had been assigned to the "green group" and received the following information:

You have been assigned to the green group. This is the group to which other participants belong. That means several other participants in this study have already tested the language training before.

Again, the latter information was provided to prevent participants from expecting that they would engage in any cooperative task with other group members. To make sure that participants realized the robot's group membership, the experimenter continued:

I am now going to present you the robot NAO ***which is a member of the other group****.*

Finally, the experimenter introduced NAO to all participants following the procedure in [20].

2.4 Dependent Measures

Acceptance of the robot. We assessed participants' *acceptance of the robot* by asking them to indicate on 7-point Likert scales how much they liked the robot NAO, and how willing they would be to get to know NAO more closely, to talk to NAO, as well as to purchase NAO. These four items were averaged and formed a reliable scale, Cronbach's α = .91, with higher values reflecting a greater extent of human-robot acceptance.

Anthropomorphization of the robot. Our measure of participants' anthropomorphic inferences about NAO is based on social psychological research demonstrating that people differentially ascribe primary and secondary emotions to in- and out-groups [14], [15] (see Section 1). More specifically, research has shown that social out-groups are often ascribed less secondary emotions (emotions that represent human essence) than the in-group, whereas there are no differences in the ascription of primary emotions. In a theoretical inversion, we used this distinction between primary and secondary emotions to measure the extent to which participants anthropomorphized the robot NAO depending on experimental condition.

As anthropomorphism often happens unconsciously [19], we applied an unobtrusive reaction-time based measure of anthropomorphism similar to a semantic priming task [22]. In this procedure, participants were presented with priming stimuli and were asked to make a decision about a stimulus word that followed. The critical stimuli represented 10 primary and 10 secondary emotion terms, whereas 20 non-emotion terms served as distractors (see Appendix). Participants were instructed to decide as quickly as possible whether the word that appeared on the computer screen represented an emotion or not, and they did so by key press. We used a 21 x 32 cm colored picture of the robot NAO as the priming stimulus and a 21 x 32 cm colored picture of a DELL computer as the control stimulus (see Figure 1). The dependent measure in this procedure was the reaction time (RT) for primary and secondary emotion words. The primary measure of anthropomorphism represented the degree to which participants responded faster to secondary emotion terms that followed the robot prime than the control stimulus.

Ten practice trials were followed by 80 test trials that were presented in random order. Both emotion stimuli and distractors were presented twice, once following the NAO prime and once following the control stimulus, respectively. On every trial, a blank screen appeared for 2000 ms followed by a fixation cross for 400 ms. Then a mask was presented for 100 ms, immediately followed by the prime picture. The prime or control stimulus appeared on the screen for 150 ms and was then again masked by an oval figure that fully covered the area of the screen occupied by the respective prime or control stimulus. Subsequently, the test word (a word representing either a primary or a secondary emotion or one of the non-emotion control words) was presented until the participant pressed one of the two decision keys.

Fig. 1. Pictures of the robot NAO and a DELL computer used as stimuli in the semantic priming task

Willingness to interact with robots in general. To measure generalization effects of the experimental manipulation on attitudes toward robots in general, participants were asked to indicate on 7-point Likert scales how willing they would be to interact with robots within nine different social contexts (e.g., in an administrative agency, as a personal fitness trainer, in health care). The nine items were averaged and formed a reliable index of *willingness to interact with robots*, Cronbach's $\alpha = .87$.

3 Results

In order to test our experimental hypotheses, we conducted a series of one-tailed *t*-tests and mixed models analyses of variance (ANOVA).

3.1 Acceptance of the Robot

Participants who perceived NAO in terms of an in-group member reported higher human-robot acceptance ($M = 5.01$, $SD = 1.41$) than participants in the out-group condition ($M = 3.64$, $SD = 1.83$), $t(41) = 2.70$, $p = .01$.

3.2 Anthropomorphization of the Robot

Preparation of Data. Average RTs were derived from correct responses only. Correct responses are those where participants correctly categorized the primary and secondary emotion words as depicting an emotion. An error was scored when participants failed to recognize the emotion word as depicting an emotion. Mean RTs for primary and secondary emotion words were calculated separately for both the NAO priming stimulus and the control stimulus. Reaction times above 1000 ms were discarded (26.64 % of all responses) in order to make sure that our measure assessed the participants' spontaneous reactions to the emotion words. The overall error rate for RTs beyond 1000 ms was 5.85 %.

Results. Results of a 2 (*experimental condition*: In-group vs. out-group) x 2 (*prime*: NAO prime vs. control) x 2 (*emotion type*: Primary vs. secondary) mixed models ANOVA yielded a significant main effect of *prime*, $F(1, 37) = 4.04$, $p = .05$, $\eta^2 = .10$, indicating that RTs were shorter when emotion terms followed the NAO prime vs. the control stimulus. Furthermore, we obtained a significant main effect of *emotion type*, $F(1, 37) = 18.41$, $p < .001$, $\eta^2 = .33$. That is, participants recognized primary emotion words faster than secondary emotion words. However, these main effects were qualified by a significant three-way interaction of *prim*e, *emotion type*, and *experimental condition*, $F(1, 37) = 5.62$, $p = .02$, $\eta^2 = .13$. No other statistically significant effects were obtained, all *p*s > .07.

To inspect this pattern of results further, we conducted two separate 2 (*prime*: NAO prime vs. control) x 2 (*emotion type*: Primary vs. secondary) ANOVAs, both for the in- and the out-group condition. For participants in the *out-group condition*, only a significant main effect of *emotion type* occurred, $F(1, 21) = 9.89$, $p = .01$, $\eta^2 = .32$. That is, participants in this condition responded faster to primary emotions ($M = 749.08$, $SD = 96.22$) than to secondary emotions ($M = 777.84$, $SD = 99.42$). No other effects were statistically significant, all *p*s > .28.

In contrast, analyses for the *in-group condition* yielded a significant main effect of *prime*, $F(1, 16) = 5.07$, $p = .04$, $\eta^2 = .24$, and a significant main effect of *emotion type*, $F(1, 16) = 8.72$, $p = .01$, $\eta^2 = .35$. However, in line with our hypothesis, these main effects were qualified by a *prime* by *emotion type* interaction, $F(1, 16) = 8.72$, $p = .01$, $\eta^2 = .35$. Results of *post hoc* comparisons showed that secondary emotions were recognized faster when NAO served as the prime ($M = 705.86$, $SD = 65.99$) than when the control stimulus preceded the emotion terms ($M = 745.25$, $SD = 67.50$), $t(16) = 4.00$, $p = .001$. In contrast, RTs for primary emotions did not differ as a function of prime ($M = 701.63$, $SD = 74.03$ for the NAO prime and $M = 691.42$, $SD = 54.37$ for the control stimulus, respectively), $t(16) = 0.91$, $p = .38$. Furthermore, when the control stimulus preceded the target words RTs were significantly shorter for primary ($M = 691.42$, $SD = 54.37$) than for secondary emotion words ($M = 745.25$, $SD = 67.46$), $t(16) = 5.19$, $p < .001$, whereas when participants were primed with NAO, participants recognized primary ($M = 701.63$, $SD = 74.03$) and secondary emotion words ($M = 705.86$, $SD = 65.99$) equally fast, $t(16) = 0.28$, $p = .78$. Figure 2 illustrates the pattern of results that fully support our hypotheses.

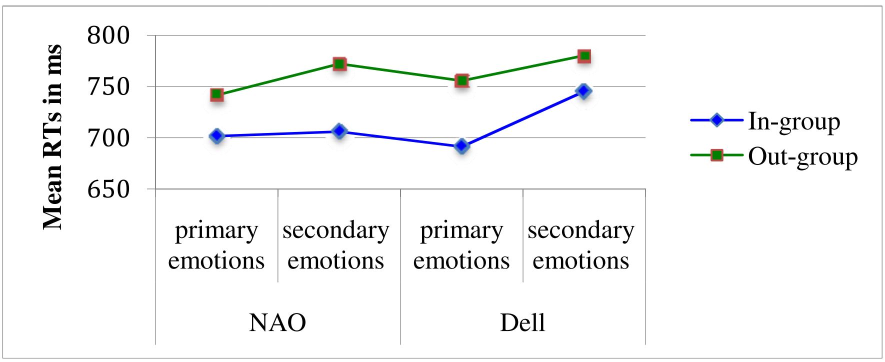


Fig. 2. Mean RTs for primary and secondary emotion words as a function of prime and experimental condition

3.3 General Willingness to Interact with Robots

We further wanted to test whether an intergroup bias toward a specific robot would transfer to robots in general. In line with our hypothesis, participants in the in-group condition indicated a greater willingness to interact with robots in every-day life ($M = 3.42$, $SD = 1.50$) compared to participants in the out-group condition ($M = 2.66$, $SD = 1.20$), $t(41) = 1.86$, $p = .04$. Accordingly, the mere idea of shared group membership with the robot NAO also affected intentions toward robots in general.

4 Discussion

Can we improve evaluations of a robot and increase people's general willingness to interact with robots by simply framing a robot as an in-group member? According to our current findings, the answer is yes. In an experiment, we manipulated whether the humanoid robot NAO was perceived as a member of participants' in-group or as an out-group member. When participants were randomly assigned to the same group as the robot NAO, they showed higher levels of acceptance than when participants were assigned to a group that merely consisted of fellow participants. Interestingly, in line with findings from intergroup research that showed that in-groups are commonly perceived as more human than out-groups [13], [14], participants anthropomorphized the in-group robot more strongly than the out-group robot. Most importantly in our view, this in-group favoritism was reflected in a greater willingness to interact with robots in general.

The present research extends existing work on the impact of social categorization of technical devices in several respects [16], [17], [18]: First, previous research has tested effects of either socially meaningful categories of high relevance for participants' social identity [16], [17] or has implemented interdependent tasks with the non-human target [18]. It is thus possible that the effects obtained by [16], [17], [18] were not caused by categorization of the non-human agent as an in- or out-group member alone. To shed more light on this issue, we used the minimal-group paradigm, and thereby we demonstrated that merely categorizing a robot as an in-group member can account for in-group favoritism toward robots. Second, we broadened the range of dependent variables by demonstrating that the in-group robot was not only evaluated more positively, but also anthropomorphized more strongly than the out-group robot (see also [16]). Extending [16], our study is the first that applied a reaction-time measure to assess anthropomorphism. However, it must be noted that our work to validate this measure is still in progress. Finally, we demonstrated generalization effects of in-group favoritism toward one specific robot on the general willingness to interact with robots in many different social contexts.

The present research provides a new and innovative strategy to improve people's attitudes toward robots and to increase their willingness to engage in human-robot-interactions. Previous research has already shown that implementing humanlike behaviors and emotions in robots can increase anthropomorphism and improve HRI (e.g., [5]). However, conveying knowledge from social psychological research on intergroup processes to the field of social robotics can easily complement these design approaches. Our data show that presenting robots as members of one's own group or team not only increases acceptance of a specific robot but also heightens people's general acceptance of robots in different social contexts. Thus, the advantage of the categorization approach beyond solely focusing on design issues lies in the broader impact on individuals' reactions toward robots.

Moreover, research on human-human interactions has demonstrated that people behave more positively in interactions with in-groupers than with out-group members; for instance, they cooperate more often with in-group members than with out-group members [23]. Likewise, it seems plausible that not only evaluations of robots but also HRI might be positively affected by a shared group membership of humans and robots. However, this assumption needs to be tested in future research that is ideally situated in a more ecologically valid experimental setting. Additionally, future research needs to investigate more closely the advantages but also possible perils of anthropomorphizing nonhumans. Although initial findings suggest positive effects on

HRI [2, 5], there is a lack of theorizing and clear empirical evidence for the impact of anthropomorphism on HRI.

Taken together, our approach seems to be a fruitful avenue for future research in the field of social robotics. Although social psychology predominantly focuses on the negative effects of group categorization, such as prejudice, discrimination and dehumanization [10], [14], [15], with the present research, we demonstrated that we could utilize such social-cognitive processes to generate positive outcomes for the acceptance of robots in every-day life.

References

1. Epley, N., Waytz, A., Cacioppo, J.T.: On Seeing Human: A Three-Factor Theory of Anthropomorphism. Psychol. Rev. 114, 864–886 (2007)
2. Luczak, H., Rötting, M., Schmidt, L.: Let's Talk: Anthropomorphization as a Means to Cope with Stress of Interacting with Technical Devices. Ergonomics 46, 1361–1374 (2003)
3. Eyssel, F., Hegel, F.: (S)he's got the Look: Gender Stereotyping of Social Robots. J. Appl. Soc. Psychol. (accepted)
4. Hegel, F., Eyssel, F., Wrede, B.: The Social Robot Flobi: Key Concepts of Industrial Design. In: Proceedings of the 19th IEEE International Symposium in Robot and Human Interactive Communication (RO-MAN 2010), pp. 120–125 (2010)
5. Eyssel, F., Hegel, F., Horstmann, G., Wagner, C.: Anthropomorphic Inferences from Emotional Nonverbal Cues: A Case Study. In: Proceedings of the 19th IEEE International Symposium in Robot and Human Interactive Communication (RO-MAN 2010), pp. 681–686 (2010)
6. Duffy, B.R.: Anthropomorphism and the Social Robot. Robot. Auton. Syst. 42, 170–190 (2003)
7. Duffy, B.R.: Fundamental Issues in Affective Intelligent Social Machines. Open Artif. Int. J. 2, 21–34 (2008)
8. Mori, M.: The Buddha in the Robot. Charles E. Tuttle Co., Japan (1982)
9. Tajfel, H., Billig, M.G., Bundy, R.P., Flament, C.: Social Categorization and Intergroup Behavior. Eur. J. Soc. Psychol. 1, 149–178 (1971)
10. Tajfel, H., Turner, J.C.: An Integrative Theory of Intergroup Conflict. In: Austin, W.G., Worchel, S. (eds.) The Social Psychology of Intergroup Relations, pp. 33–47. Brooks/Cole, Oxford, England (1979)
11. Gaertner, S.L., Dovidio, J.F.: Reducing Intergroup Bias: The Common Ingroup Identity Model. Psychol. Press, New York (2000)
12. Jackson, L.A., Hodge, C.N., Gerard, D.A., Ingram, J.M., Ervin, K.S., Sheppard, L.A.: Cognition, Affect, and Behavior in the Prediction of Group Attitudes. Pers. Soc. Psychol. B. 22, 306–316 (1996)
13. Haslam, N.: Dehumanization: An Integrative Review. Pers. Soc. Psychol. Rev. 10, 252–264 (2006)
14. Leyens, J.-P., Paladino, M.P., Rodríguez-Torres, R., Vaes, J., Demoulin, S., Rodríguez-Pérez, A., et al.: The Emotional Side of Prejudice: The Attribution of Secondary Emotions to Ingroups and Outgroups. Pers. Soc. Psychol. Rev. 4, 186–197 (2000)
15. Leyens, J.-P., Rodriguez, A.P., Rodriguez, R.T., Gaunt, R., Paladino, P.M., Vaes, J., Demoulin, S.: Psychological Essentialism and the Attribution of Uniquely Human Emotions to Ingroups and Outgroups. Eur. J. Soc. Psychol. 31, 395–411 (2001)

16. Eyssel, F., Kuchenbrandt, D.: My Robot is More Human than Yours: Effects of Group Membership on Anthropomorphic Judgments of the Social Robot Flobi. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2011) (accepted)
17. Nass, C., Isbister, K., Lee, E.-J.: Truth is Beauty: Researching Embodied Conversational Agents. In: Cassells, J. (ed.) Embodied Conversational Agents, pp. 374–402. MIT Press, Cambridge (2000)
18. Nass, C., Fogg, B.J., Moon, Y.: Can Computers be Teammates? Int. J. Hum.-Comput. Int. 45, 669–678 (1996)
19. Nass, C., Moon, Y.: Machines and mindlessness: Social Responses to Computers. J. Soc. Issues 56, 81–103 (2000)
20. Pierris, G., Dahl, T.S.: Compressed Sparse Code Hierarchical SOM on Learning and Reproducing Gestures in Humanoid Robots. In: Proceedings of the 19th IEEE International Symposium in Robot and Human Interactive Communication (RO-MAN 2010), pp. 681–686 (2010)
21. Eyssel, F., Kuchenbrandt, D.: Manipulating Anthropomorphic Inferences about NAO: The Role of Situational and Dispositional Aspects of Effectance Motivation. In: Proceedings of the 20th IEEE International Symposium in Robot and Human Interactive Communication (RO-MAN 2011), pp. 467–472 (2011)
22. Wittenbrink, B., Judd, C.M., Park, B.: Evaluative versus Conceptual Judgments in Automatic Stereotyping and Prejudice. J. Exp. Soc. Psychol. 37, 244–252 (2001)
23. Goette, L., Huffman, D., Meier, S.: The Impact of Group Membership on Cooperation and Norm Enforcement: Evidence using Random Assignment to Real Social Groups. The Am. Ec. Rev. 96, 212–216 (2006)

Appendix

Primary emotions: Excitement, joy, surprise, happiness, pleasure, anxiety, fear, pain, sadness, anger

Secondary emotions: Love, hope, passion, emotion, admiration, contempt, guilt, shame, bitterness, spitefulness

Distractor Words: Car, tree, concrete, letterpress, sidewalk, butter, ticket, shutter, figure, terrain, autumn, mall, pebble, lampshades, lineal, mobile phone, stake, postcard, sailboat, track, door, bird, water, fence

Homewrecker 2.0: An Exploration of Liability for Heart Balm Torts Involving AI Humanoid Consorts

Sonya Ziaja

Oxford University, Wolfson College, Oxford, UK
sonya.ziaja@wolfson.ox.ac.uk

Abstract. With the development of artificially intelligent humanoid consorts, robotics is venturing into a realm of legal liability that has traditionally governed social interactions between humans and other humans, rather than interactions between humans and machines. How can and should legal systems deal with the problems that arise in regulated human interpersonal and sexual relationships when there is an AI sex doll in the mix? Heart balm torts, traditionally used to hold a third party paramour civilly liable for the dissolution of a protected relationship, provide a potential answer. Finding an appropriate entity to be liable will be problematic, though robot producers and the AI entity itself could both be potential defendants in a heart balm case. Producers may be able to limit liability if they can incorporate the experience of heartbreak and compassion into their creations.

Keywords: Liability, Social Robots, Artificial Intelligence, Sex, Heart Balm Tort.

1 Introduction

Rapid development and interest in introducing artificially intelligent ("AI") robots into daily human life necessitates taking a fresh look at the ability of our existing legal systems to handle cases involving artificially intelligent actors. Robot manufacturing and design has long faced the prospect of product liability suits. Now, however, we are developing social robots that are designed to interact with humans in intimate ways when we are at our most human and vulnerable. With this shift in purpose, robotics is venturing into a realm of legal liability that has traditionally governed social interactions between humans, rather than interactions between humans and machines.

New purposes for service robots—military [1], sentry [2], care of the ill and elderly [3], companionship and sex [4]—each presents new challenges for the legal systems of the world and highlights unresolved ethics issues within the law deserving of scholarship. This paper concentrates on one type of social robot—the "sexbot" or romantic companion—and the civil law that relates to sexual and romantic relationships—the heart balm torts.

In his 2007 book, David Levy predicted that "[l]ove and sex with robots on a grand scale are inevitable." [4] In some ways, the AI sexbot is already an established technology. Humanoid sexbots with multiple choices of "personalities" are already on the

B. Mutlu et al. (Eds.): ICSR 2011, LNAI 7072, pp. 114–124, 2011.

market. What is not established though is regulation and law that will be able to adroitly handle conflicts that will inevitably arise between humans and AI sex dolls.

The entry of such dolls into the market presents new questions for legal systems. While it is easy to forecast, as much science fiction has, numerous civil and criminal causes of action that will challenge the legal systems of the world, this article will focus only on an overlooked and basic problem—how can and should law deal with the problems that arise in regulated human interpersonal relationships when there is an AI sex doll in the mix? For example, if a doll's owner becomes enamored with the doll, and leaves his spouse, can the spouse sue as she or he would be able to if the interloper had been human? And, who in this situation would be sued? The manufacturer? The inventor? Or, perhaps the AI itself?

This article gives the necessary background to explore these questions—describing first the current state of AI consorts available on the market, and presenting an overview of the heart balm torts. This article will then examine who should be held liable if an AI entity committed the tort. Finally, this article posits how law and robotics may be able to work together to protect human values while limiting liability.

2 Background

As an initial matter, artificial intelligence as it is used in this article means only that the robot has sufficient responsive and interactive capabilities to make people feel as though the robot is intelligent. In order to discuss the legal ramifications of AI robots in intimate human relationships, the Turing test approach to AI suggested by Professor Lawrence Solum [6-7] is sufficient. This paper favors this definition of AI, because in a court the actual mental state (mens rea) of a defendant does not matter as much as what the people judging the defendant—the judge or jury—believe the defendant's state of mind to be. By using the word robot, this article takes a broad definition, meaning a constructed machine that responds to external stimuli and is designed with a specific purpose to serve humans.

2.1 Technology

Humans have been developing technologies for sexual purposes for thousands of years.[1] Our recent endeavor has been to create a fully automated humanoid robot that is responsive and has a personality. Social companion robot projects around the world have produced AI robots that are meant to be integrated into daily human affairs. South Korea has set a goal to have a robot in every home within the next decade. [8]

---

[1] The use of inanimate objects as sex aids is not something new to humanity. Only recently the world's "oldest sex toy" (dating back 28,000 years) was found by archeologists from the University of Tubingen in Germany. See Rosemary Black, Prehistoric siltstone phallus, the world's oldest sex toy, was also used as tool to ignite fires, NY Daily News (May 17, 2010), available at http://www.nydailynews.com/lifestyle/2010/05/17/20100517_prehistoric_siltstone_phallus_the_worlds_oldest_sex_toy_was_also_used_as_tool_to.html

Among the many projects currently in development or on the market—including Asimo [9], Olivia [10], the robot seal Paro [11], or the substitute boyfriend Primo Puel [12]—TrueCompanion's sex robot doll, Roxxxy, has dominated public media. [13-14] Taken together, these robots indicate that AI robots are already active parts of daily human life.

While there are many social companion robots in the works with sophisticated interactive capabilities than Roxxxy, this article will focus on a robot with Roxxxy's capabilities. This is primarily because the Roxxxy robot is made explicitly for both sex and companionship, and so is most likely to be involved in the regulated area of human sexual relationships.[2] Further, the technological attributes and purposes of this AI sex doll are a sufficient catalyst to explore how conflicts between AI consorts and humans might be dealt with in legal systems.

Roxxxy was unveiled in 2010 at a Las Vegas Adult Entertainment Expo. [14] Physically, the doll is constructed to look and feel like a woman and is anatomically correct in all of the areas that matter most for its purpose. [15] It has a "heartbeat" and circulatory system that distributes and expels heat from its body. [15] It also snores. [16]

For the communication part of the doll's purpose, it responds to both touch and what people say to it. [15] It remembers preferences, can send fawning emails while the owner is away, and hold conversations about seemingly anything. [17] The doll comes with five preset "personalities." [15] It can act sexually adventurous, frigid, youthful, coy, or motherly, depending on its setting. [15] Users can also customize the doll's personality and swap personalities with other users. [15] The TrueCompanion website advertises the ability to "build your own additional Girlfriend Personalities." [15]

The doll is an attempt to put a complete personality[3] in a humanoid robot. [18] It does have room to become more complex and human-like. But, its current qualities are already sufficient for it to become involved in human relationship conflicts where there is a legal remedy.

2.2 The Heart Balm Torts

Certain relationships are highly valued by society and offered protection through the law. Among these are intimate familial relationships between parents and children and between spouses. Tort laws developed over time to protect these relationships from being damaged or destroyed by third parties. These include the historical torts of enticement and seduction and their modern equivalents—alienation of affections and

---

[2] Additionally, according to the inventor there have already been thousands of preorders for the $7000 doll. Brandon Griggs, Inventor unveils $7,000 talking sex robot, CNN (February 1, 2010).

[3] Douglas Hines stated that the idea for creating Roxxxy began with the desire to preserve human personalities of friends and loved ones.

criminal conversion.[4] Because of these torts are concerned with preventing and mending heartbreak, they are called amatory or heart balm torts.

To make a successful claim for a alienation of affections, traditionally, a plaintiff would have to prove three elements: (1) that true affection had existed between the spouses at one time; (2) that the affection was destroyed; and (3) that the defendant caused the destruction of affection or otherwise impaired the marital relationship. [19] To cause the destruction of marital affection does not mean that the defendant had to be the sole cause. Rather the defendant had to be a substantial factor in the deterioration of the affection between the spouses. [20] The plaintiff also does not have to prove that the defendant had malicious intent or other improper motives to be successful on a claim for alienation of affections. The defendant needs only to have acted without justification. [20] Moreover, the defendant could be liable even if he or she was not a sexual partner of the plaintiff's spouse. [20] In-laws, ministers, employers, and churches have all been defendants in previous alienation of affections cases. [20] A relationship that interferes with marital affection can be completely platonic, and the interloper can still be liable for alienation of affections. The only other limiting factor for this tort is that the defendant has to have had knowledge of the marriage in order to be liable. [21]

Criminal conversion differs from alienation of affections in that it focuses solely on punishing and preventing adulterous relationships. [21] In these cases, therefore, the plaintiff needs to prove that there was a sexual relationship between the defendant and the plaintiff's spouse. Unlike alienation of affections, ignorance of the marital relationship is not a defense for criminal conversion. In practice, criminal conversion is said to resemble a strict liability tort, because there are so few defenses for a defendant. [22]

The heart balm torts and the basis for heart balm claims have evolved—and in some cases, died—along with changes in cultures. The original basis for heart balm claims was based in property law and the idea that a man had a property interest in his wife. [23-24] This meant that only men could sue for alienation of affection and criminal conversion; the property right was a justification for monetary compensation. [24] But as people grew to recognize the rights of women and mutual responsibility for relationships, the torts changed. With the passage of the 1844 Married Women's Acts, women, like men, could recover for “interferences with [their] spousal relations.” [19] The basis for maintaining the heart balm torts became protecting marriage from interference rather than a cold property issue. [25-26]

In past half-century, the torts have continued to change. Because defending against heart balm torts is so difficult, people began to use the torts to extort money, forcing possible defendants to settle to avoid litigation. [20], [24] Over time, concern about extortion and a rising belief that spouses were mutually responsible for their marriage led many jurisdictions to abolish the torts through legislation and judicial action. [24] Nonetheless, the heart balm torts persist. And even where they have been abolished, it has not stopped the flow of adultery-based claims in court. [20]

---

[4] The heart balm torts also include a right to sue for a breach of promise to marry; but discussion of this tort with relation to social robots is beyond the scope of this article.

Courts have generally seen two separate rationales for the existence of heart balm torts—property interest of the husband and preserving the marriage. In the twentieth century, jurisdictions rejected the first rationale as outdated. The goal behind the second rationale, however, that marriages should be protected from interference, has not been rejected. [27] This is true even within the same court opinions that judicially abolished the torts in some states. [28] Rather judges found that there was no evidence that the heart balm torts actually worked to effectuate that purpose. A Missouri court abolishing alienation of affections summarized the common argument nicely, "the original property concepts justifying the tort are inconsistent with modern law [and] the modern justification that the tort preserves marriages is a fiction." [29] New Jersey's Supreme Court also weighed in, abolishing the tort because "no preventative purpose is served, since such torts seldom are committed with deliberate plan." [28] These cases suggest that if social circumstances changed, for example due to the introduction of artificially intelligent romantic companions, such that the heart balm torts could protect marriages, there would be room for heart balm torts to return to those jurisdictions.

There is a third explanation, which is that it is the pain of heartbreak, in the context of a protected relationship [27], that society looks to compensate through the heart balm torts. In cases of adultery, spouses can suffer emotional damage, loss of trust, mental insecurities, decreased job performance, and financial insecurity. [20] Indeed, medical researchers have found that there is real physical pain experienced through heartbreak. [30] This explanation for the existence of the torts is consistent with the continuation of adultery laws and cases [27] and with the continued taboo against breaking up marriages. This theory too leaves open the possibility of liability for AI sexbots.

The Louisiana Supreme Court in 1928, opining about why there was no recognition of heart balm torts in Louisiana's jurisprudence wrote that "the law undertakes only to control and regulate human conduct—not human nature." [31] A robot involved in a heart balm tort, however, does not have a human nature. It only has conduct, unless it is programmed to have a full range of intelligent emotional responses.

One might consider the possibility that other non-agents, besides robots, might be defendants in heart balm cases. There are few documented cases of non-humans being sued for heart balm torts. [31] This may be in part because many of these suits settle outside of court, as they are famously difficult to defend. Even so, AI sexbots are distinguishable from other technologies that have been blamed for destroying marriages. For example, online gaming addiction was recently cited in a UK survey as a significant reason for divorce. [33] But, generally speaking people do not view games as responsible individuals. Whereas we do seem to have that expectation for robots that mimic pets and people. Timothy Bickmore's 1998 paper on "Friendship and Intimacy in the Digital Age" suggests that robotic friends might be a solution for those people who do not have close friendships. [34] Friendships and relationships with robots work as friendships though because people feel that the robots have agency, or as Gail Melson puts it, "moral standing." [35], [5] The difference between seeing a machine as only a tool (as in the case with online gaming) versus trusting it as a moral agent or even feeling like it is a moral agent is what sets robot companions apart from other non-agents.

3 Liability Options

If a heart balm suit were to be brought by a spouse where the third party was a sexbot instead of another human, who would be liable? The heart balm torts are directed against the third party paramour. In this hypothetical though, the lover is not human, nor a legal person.[5] When presented with a case like this, legal systems will have to respond by either finding a human at fault (the inventor, manufacturer, the owner), or by finding the sexbot at fault. Either of these options is problematic.

Placing liability with a human or legal person does have its benefits. Legal persons can be sued. They (mostly) have the capacity to defend themselves and hire attorneys. And, perhaps most importantly, they have money and assets that can be used to pay damages in case the plaintiff wins, or settle before trial. The problem then becomes which persons should be liable?

The inventor could be liable because the doll is the creation of the inventor. And if the plaintiff needs to demonstrate the intent of the defendant, it is a relatively simple matter to demonstrate that inventors of AI consorts intended to create a robot that would be a “true” companion. Indeed, the inventor has an interest in selling the robots, and is likely to make them as appealing as possible. And, with ample research and media reports on technology addiction, it should be foreseeable to inventors that a well-designed sexbot might be a problem. Moreover, for heart balm torts, no specific malicious intent is necessary to be liable. For criminal conversion, no intent is necessary, so long as the interloper had sex with a married person.

For something as complex as an AI humanoid consort, there are multiple inventors, engineers, psychologists, make-up experts and models involved. No one of these could reasonably be singled out as being at fault. It is far more reasonable and likely that the production company would hold the liability.

The manufacturer of the doll is another possibility for liability. There is a significant history of manufacturer liability for products that cause harm to people. [36] However, for the manufacturer to be liable under a products liability theory, the product must be deficient in some way. Here, though, the issue is not that an AI sex doll was physically dangerous and injured someone, but rather it was designed well enough to draw its owner’s lust and companionship away from his or her spouse and direct it towards the doll.

One might be tempted to place the blame on the owner of the AI doll. Depending on the technology used, it is the owner who ultimately determines the doll’s personality and physical features. [15] In effect, the owner programs the doll to help him (or her) realize his fantasies. From this perspective, it is ultimately the owner’s “fault” that he is so drawn to the doll. But, this argument falls apart quickly. Liability

[5] Legal persons do not have to be human. An artificial person, according to Ballentine’s Law Dictionary, as of 2010, is “A person created by law or by authority of law, such as a corporation, as distinguished from a natural person, that is, a human being.”

does not make sense to rest in the owner for simple economic reasons. Property in marriage is shared, and so any damages that a plaintiff might be able to get from the owner, would also have been the plaintiff's assets.

While some jurisdictions may find liability for heart balm torts with persons involved in creating the sexbot, some scholars and transhumanists support creating legal personhood for AI robots. [6] Without legal personhood, it is difficult to imagine that an AI could be brought to court, let alone be liable for a heart balm tort. Legal personhood does not have anything to do, though, with being a non-human.

Indeed if an AI were sued for alienation of affections it would not be the first time a non-human agent was brought to court for a heart balm tort. In 1987, a man in Idaho sued his wife's church for alienation of affections. [31] The woman's church, along with her mother, convinced her that her marriage was not sanctioned by the church and therefore a sinful relationship which she needed to distance herself from. The man's case failed, but not because a named defendant was not human. [31]

There is no set rule for establishing legal personhood or quasi-personhood for a non-human entity. [37] For example, while a corporation is a legal person in both France and England, both countries arrived at that conclusion through a different process. [37] In France, it was legislated, and has roots in Roman Law. [37] Whereas in England, the conclusion came about in a more organic way; scholars are still debating how and why corporations have a legal person status. [37]

Scholars have suggested a variety of roadmaps to determine the validity of granting legal personhood for AI entities. [6] Professor Solum, in his foundational article on legal personhood for artificially intelligent actors, suggests that the path to legal personhood begins with an inquiry as to "whether or not the entity can and should be made the subject of a set of legal rights and duties." [6] Solum answers the ability prong of this inquiry by using the Turing test—in essence, if a human cannot tell it is a robot, then it might as well be human.[6] [6] Other scholars, however, argue that intentionality is the key component to legal persons. [38-39]

Even if an AI entity is given legal personhood, in heart balm cases there are enormous hurdles to successfully suing the AI humanoid. So much so that it is simply not a practical solution currently. Because of the nature of AI robots, there is no means to effectuate the purpose of the right of action. The economics of suing the AI itself does not make sense. One of the purposes of the tort law system is to restore the injured party to her position before the injury occurred. [40] Setting aside questions about the validity of a system that puts a monetary value on affection, sex and companionship, the plaintiff in any of the heart balm torts is compensated with money damages. [24] Unless the AI sexbot is able to earn money of its own, it presumably does not have any assets of its own that could be awarded to the plaintiff.[7]

Moreover, it is questionable whether the AI would have the capacity to be sued at all. Capacity in this sense means "[the] ability to understand the nature and effect of

[6] An advantage of this method is that one does not need to delve into philosophical questions about intention and intelligence.

[7] A guilty defendant sexbot could in theory be sold to pay the plaintiff. This brings its own ethical challenges about whether or not such a sale would be slavery, whether renting an AI robot for sex is criminal pimping, as well as a host of other questions about the scope of "rights" of AI.

the act in which he is engaged and the business which he is transacting." [41] In essence, in order to have capacity, the AI would need to be self-aware.[8] Here again, however, a Turing test approach to capacity might suggest that an AI sexbot does have sufficient understanding to be sued.

In order to effectuate the goals of the heart balm torts—compensating the plaintiff and preserving protected relationships—liability should be held by the entity that has the ability to change or control the sexbot's behavior and has the ability to compensate the aggrieved plaintiff. The inventor-producer and the AI entity seem to be the most likely candidates for control over the AI's behavior. For now, however, because an AI entity does not yet have a means to compensate a plaintiff, liability will most likely rest with the inventor-producers.

Inventor-producers may face a variety of other lawsuits related to heart balm torts as well. For example in jurisdictions that had previously abolished the tort, courts might find that the underlying duty—to not unreasonably interfere and break up marriages—still exists. If that duty is present, courts may consider creating a hybrid tort, somewhere between product liability, negligence, and the heart balm torts.

4 Incorporating Heartache as a Means to Limit Liability

The heart balm torts were a way of taking what human cultures already valued—that it is unjust to break apart a married couple—and codifying it in law. The torts have evolved, however, over time as people began to feel more comfortable relying on social rather than legal pressure to protect relationships. Social pressure works because the human capacity for empathy and compassion allows for that trust. Having known pain of heart break, one naturally does not want another person to have to suffer through that pain and generally does not want to be the cause of it. In other words, people rely on one another for moral feedback.[9]

In the case of sexbots, people could easily put a similar social trust in them, even while the bots might not have the same ability to respond as we expect humans to. Humans already hold this expectation for robots that are less intimate and less human-looking than a sexbot would be. According to researches Nass and Moon, people already "apply social rules and expectations to computers." [43] In the case AIBO, which Sony bills as being "a true companion with real emotions and instinct", the robot pet is designed to "encourage owners to project humanlike attributes onto [them]." [5] Furthermore, in a survey of AIBO users, 76% said that AIBO had "moral

---

[8] Whether it would be desirable, let alone possible, to create self-aware AI is well beyond the scope of this article. But it is safe to say that self-aware AI would open a pandora's box of ethical problems for human society. As Professor Jos De Mul puts it, "just as the ape cannot form an adequate picture of the human life form, so it is not given to us to visualize the nature or attractiveness of these new life forms." [42]

[9] Sometimes, of course, this feedback is flawed and at other times might not happen at all. Still, in general, people depend on each other to keep one another "honest." One can see this principle employed in AA type groups that encourage members to find new groups of friends and distance themselves from people who might encourage them to engage in self-destructive behaviors.

standing"; or as David Levy put it "it could be held morally blameworthy for its actions and could have rights and deserve respect." [5], [35] If people are willing to ascribe moral standing a robotic pet, people are probably willing to ascribe moral standing to a robotic humanoid companion.

If an AI sexbot is not able to provide the moral feedback that people expect of one another, then the preventative goal of the heart balm torts dictates that inventor-producers should be held liable. While the heart balm torts may not have the effect of being preventative when only humans are involved, they could be preventative when robots are. Holding inventor-producers liable means that they would internalize—in monetary form—the social damage caused by robots. This damage, or even the threat of monetary damage, could pressure inventor-producers to alter the design of robots so that they comport with social standards that people expect of one another.

One potential way to design robots to meet this standard is to introduce the feeling of heartbreak along with care for other people in its owner's life. If the robot can consider the emotions not only of its owner—but of the owner's spouse, children, etc.—and can "know" and avoid those actions that might harm those protected relationships, then the robot might meet the standard that we expect our friends, relatives and lovers to conform to. If the robot is designed to meet this standard, then the producer-inventor should be considered to have done its due diligence before releasing a new social entity into society, and should therefore not be held liable for the actions of the AI robot.

The idea of creating emotions for robots is not new. A recent example, Kim Jong-Hwan's "robot chromosomes" are "intended to give ability to [robots to] reason and to feel desire and lust, just like us." [44], [5] In order for robots to enter into human romantic relationships in a way that is consistent with the values underlying the heart balm torts, it may also need to experience heartache and empathy as we do.

5 Conclusion

The inclusion of AI interactive social robots into human relationships will lead to liability previously only encountered in human-human relationships. In order to protect human values, law and robotics will have to adapt to each other. Jurisdictions can revive the heart balm torts, combined with product liability and negligence theory in order to pressure inventor-producers to create humanoid consorts, sexbots, etc. that meet the social standards that we expect of other people. One key characteristic of emotionally responsible human relationships is that we are able to feel heartbreak as well as compassion and empathy. To create emotionally and legally responsible robots, social robot designers should instill the wisdom gained through heartbreak into their creations.

Acknowledgements. This paper benefitted from the advice and generosity of many people, including Professor Jaime King at the University of California-Hastings College of the Law, L.F., and H.G., who were willing to indulge playful hypotheticals. Thank you all. Most of all, thank you to Andrew Ziaja for his support, love, and triple-stitched wisdom. This would not have been possible without you.

References

1. Arkin, R.: ACM/IEEE International Conference on Human-Robot Interaction. In: Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction (2008), http://portal.acm.org
2. Lee, Y.: Samsung Gets Creative. Bloomberg Markets Magazine (presenting Samsung's machine-gun-equipped sentry robot) (August 31, 2010), http://www.bloomberg.com/news/2010-08-31/samsung-gets-creative.html
3. Aging Japan Building Robots To Look After Elderly, Terra Daily, March 15 (2006), http://www.terradaily.com/reports/Aging_Japan_Building_Robots_To_Look_After_Elderly.html
4. Levy, D.: Love and Sex with Robots. Harpers (2007)
5. Solum, L.B.: Legal Personhood for Artificial Intelligence. North Carolina Law Review 70 (1992)
6. Asaro, P.: Robots and Responsibility from a Legal Perspective. In: Proceedings of the IEEE Conference on Robotics and Automation, Rome, April 14 (2007)
7. Lovgren, S.: A Robot in Every Home by 2020, South Korea Says, National Geographic News, September 06 (2006)
8. The Intelligent Asimo: Systems Overview and Integration. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, vol. 3, pp. 2478–2483 (2002)
9. Guizzo, E.: Singapore Researchers Unveil Social Robot Olivia. IEEE spectrum, August 13 (2010), http://spectrum.IEEE.org/automaton/robotics/home-robots/social-robot-olivia
10. Wada, et al.: Living With Seal Robots—Its Sociopsychological and Physiological Influences on the Elderly at a Care House. IEEE Transactions on Robotics 23(5), 972–980
11. Sharkley & Sharkley: Granny and the robots: ethical issues in robot care for the elderly. Ethics and Information Technology, July 3 (2010)
12. Sexrobot, I.: Harpers Magazine (March 2010)
13. Karlin, S.: Red-Hot Robots, Roxxxy and Rocky, the world's first sex robots, are ready to leave the lab. IEEE Spectrum (June 2010), http://spectrum.IEEE.org/robotics/humanoids/redhot-robots
14. Frequently Asked Questions, TrueCompanion.com (last checked September 14, 2010), http://www.truecompanion.com/shop/pages/FAQ.html
15. Griggs, B.: Inventor unveils $7,000 talking sex robot, CNN (February 1 (2010)
16. Hornyak, R.: Sex Robot Roxxxy Looking for Action, cnet news, January 11 (remarking on the doll's ability to discuss Manchester United) (2010), http://news.cnet.com/8301-17938_105-10432597-1.html
17. Chapman, G.: World's first life-size robotic girlfriend, AFP, January 10 (2010), http://www.news.com.au/technology/worlds-first-life-size-robotic-girlfriend/story-e6frfro0-1225817805297
18. Restatement (Second) of Torts, 684 (1977)
19. Patel, N.: Note, The State's Perpetual Protection of Adultery: Examining Koester v. Pollard and Wisconsin's Faded Adultery Torts. Wisc. L. Rev. 1013 (2003)
20. Page Keeton, W., et al.: Prosser and Keaton on the Law of Torts, 5th edn., vol. 920 (1984)
21. Kavanagh, Note: Alienation of Affections and Criminal Conversion: Unholy Marriage in Need of Annulment, 23 Ariz. L. Rev. 323, 324 (1981)
22. Prosser and Keaton, The Law of Torts, 2nd edn., vol. 697 (1955)

23. Heard, J.: Comment: The National Trend of Abolishing Actions for the Alienation of a Spouse's Affection and Mississippi's Refusal to Follow Suit, 28 Miss. C. L. Rev. 313, 316–319 (2009)
24. Corbett, W.R.: A Somewhat Modest Proposal to Prevent Adultery and Save Families: Two Old Torts Looking for a New Career, 33 Ariz. St. L.J. 985 (2010)
25. Deans, S.M.: The Forgotten Side of the Battlefield in America's War on Infidelity: A Call for the Revamping, Reviving, and Reworking of Criminal Conversation and Alienation of Affections, 53 How. L.J. 377 (2010)
26. Nicolas, P.: The Lavender Letter: Applying the Law of Adultery to Same Sex Couples and Same Sex Conduct, 63 Fla. L. Rev. 97 (2011)
27. Magrierowski v. Buckley, 121 A.2d 749 (N.J. Super. 1956)
28. Helsel v. Noellsch, 107 S.W.3d 231 (Supreme Court of Missouri 2003)
29. Kross, et al.: Social Rejection shares somatosensory representations with physical pain. Proceedings of the National Academy of Sciences 108(15), 6270–6275 (2011)
30. O'Neil v. Schuckardt, 733 P.2d 693 (Idaho 1987)
31. Moulin v. Monteleone, 115 So. 447, 456 (La. 1928)
32. Divorce Online Blog, June 1 (2011), `http://blog.divorce-online.co.uk/?p=2014`
33. Bickmore, T.: Friendship and Intimacy in the Digital Age (1998)
34. Melson, G.: Robotic Pets in Human Lives: Implications for the Human-Animal Bond and for Human Relationships with Personified Technologies. Journal of Social Issues 65(3), 545–567 (2009)
35. Hackney Jr., J.R.: Forum: On the Intellectual History of Law and Economics: Law and Neoclassical Economics: Science, Politics, and the Reconfiguration of American Tort Law Theory. 15 Law & Hist. Rev. 275 (1997)
36. Foster, N.H.D.: Company Law Theory in Comparative Perspective: England and France, 48 Am. J. Comp. L. 573 (2000)
37. Gray, H.C.: The Nature and Sources of the Law (Roland Gray ed., MacMillan 1921) (1909)
38. Searle, J.: Minds, brains and programs. Behavioral and Brain Sciences 3, 417–457 (1980)
39. Theories of Tort Law, Stanford Encyclopedia of Philosophy, `http://plato.stanford.edu/entries/tort-theories`
40. Balantine's Legal Dictionary, Copyright (c) 2010 LexisNexis
41. De Mul, J.: Transhumanisme - de convergentie van evolutie, humanisme en informatietechnologie. Tydskrif vir Geesteswetenskappe 45(2), 159–176, available in English at `http://www.xs4all.nl/~rekveld/tech/DeMul_Transhumanism.pdf`
42. Nass, Moon: Machines and Mindlessness: Social Responses to Computers. Journal of Social Issues Spring (2000)
43. Kim, J.-H., et al.: The Origin of the Artificial Species: Humanoid RoborHanSaRam,Robot Intelligence Laboratory EECS Department, KAIST, Guseong-dong, Yuseong-gu, Daejeon 305-701 Republic of Korea

Examining the Frankenstein Syndrome

An Open-Ended Cross-Cultural Survey

Dag Sverre Syrdal[1], Tatsuya Nomura[2], Hiroto Hirai[2], and Kerstin Dautenhahn[1]

[1]Adaptive Systems Research Group, School of Computer Science,
University of Hertfordshire, Hatfield, UK
{d.s.syrdal,k.dautenhahn}@herts.ac.uk
[2] Department of Media Informatics, Faculty of Science and Technology,
Ryukoku University, Japan
{nomura@rins,t070454@mail}.ryukoku.ac.jp

Abstract. This paper reports findings from an open-ended survey on attitudes towards humanoid robots collected from samples in the United Kingdom and Japan. 335 participants were asked how they felt about humanoid robots becoming widespread in society and what tasks they wanted humanoid robots to perform. While the UK sample was overall less negative towards humanoid robots than their Japanese counterparts, the UK sample did not want robots to perform tasks that required capabilities deemed as human qualities, such as empathy, caring, or independent decision making.

Keywords: Survey, Humanoid Robots, Attitudes.

1 Introduction

1.1 Kaplan and the Frankenstein Syndrome

The differences between Western and Japanese attitudes and responses towards technologies have been a subject of fascination across different fields of research. Kaplan [1] argues that much of the work considering such differences have taken as the baseline that the Japanese have a greater fascination with these technologies, than their western counterparts. Kaplan, however, makes the argument that this is not necessarily the case. According to Kaplan, acts of technological creation and the use of technological creations are seen as very important in defining the identity of both creator and user in Western Cultures. As such, the role of technology in the West is fraught with many taboos. One of these taboos can be described as a Frankenstein Syndrome, wherein the creation of an artifact that is a convergence between humanity and technology is an act of potential transgression in and of itself. Many Western narratives deal with the implications of such transgressions. Mary Shelley's novel [2], may be the most iconic, but this narrative is repeated in almost every single narrative that deals with robots. From Blade Runner [3] to the recent children's movie Wall-E [4], the creation of artificial beings is seen as problematic, with potentially disastrous consequences. Japan, however, Kaplan goes on to argue, has no such strong taboos regarding such technologies, they do not occupy a special position and thus their

B. Mutlu et al. (Eds.): ICSR 2011, LNAI 7072, pp. 125–134, 2011.

creation and use are only subject to the same rules governing general conduct within society. An example of this line of thinking would be Masahiro Mori's argument in the Buddha in the Robot [5] that the robot and its operator are both acting within the same moral framework.

1.2 Cross-Cultural Differences in HRI

That differences exist between Japan and the West has, of course, been considered within the field of HRI. Attempts at exploring and quantifying these differences have been numerous [6-9]. Overall, these studies have challenged the popular belief that the Japanese are more positive towards robots than people in the West. A series of studies using the Negative Attitudes towards Robots Scale (NARS [10]) as a measure for attitudes towards robots, were performed by Bartneck et al. [6-7]. The NARS was a scale developed using a method derived from Allport's Lexical Hypothesis [11], where open-ended responses from surveys where categorized and turned into questionnaire items, responses to which were then subjected to factor analysis within large Japanese samples. When this test, along with translations was administered to samples of different cultures, Japanese samples did not score significantly more positively than Dutch or American samples, and in fact scored more negatively than Western samples on some dimensions. Syrdal et al. [12] drew attention to the possibility of cultural specificity of the NARS and similar scales constructed using this method, suggesting that even if appropriately translated, the NARS may still rely on culturally specific constructs, leading to results not being comparable across cultures. A similar issue was highlighted by MacDorman [9], who found that different measures for attitudes and responses to robots yield different results in terms of cultural differences.

1.3 Towards a Tool for Cross-Cultural Investigation of the Frankenstein Syndrome

Due to these factors, it is reasonable to suggest that tools to study the specific interactions of culture on attitudes towards humanoid robots, need to be developed cross-culturally from their inception to their deployment. Using the lexical hypothesis as base, this paper describes the initial stage of data-collection from open-ended questionnaires to a large number of participants in both the United Kingdom and Japan.

2 Methodology

2.1 Questionnaire Design

A questionnaire was created for this study, consisting of 3 questions apart from demographics. The questions were open-ended in order to get as wide a range of responses as possible from participants. Apart from demographic details, the questions were as follows:

- **Q1. How do you feel about humanoid robots becoming widespread in society?**
- **Q2. What sort of activities do you think humanoid robots should perform in society, what sort activities do you think they should not perform?**
- **Q3. Where do your impressions of humanoid robots come from?**

2.2 Data Collection

The Japanese student sample was collected through students voluntarily participating in the survey conducted at the end of their lectures. The non-student sample was collected using postal mailings to request responses from the employees of different companies. The UK sample was collected via snowball-sampling where participants who participated in other HRI experiments and filled out the survey, volunteered to give out the questionnaires to their acquaintances, colleagues, affinity groups and classmates, allowing for exposure to a more diverse population than direct recruitment would have.

2.3 Coding Scheme

A data-driven coding scheme was created and is presented below with two examples from each class for Q1 and Q2.

Q1 Positive

0 – No Positive Sentiment

1 – Expectation of specific benefits, future possibilities

"I think that they can be quite useful."

"I think I would like to have a robot to help me in the home. It could be fun to talk to."

2 – Other positive, including general positive sentiment

"It sounds promising."

"I would be happy to see it happen."

Q1 Negative

0 – No Negative Sentiment

1 – Negative changes to human mental and social processes (laziness, unemployment, meaning of humanity).

"I am afraid that their presence will encourage less human interaction and make our world more technical - less natural."

"We have managed ok so far, - it would just make us more lazy,"

2 – Other negative, including physical risks and maintenance costs and simple negative expression

"Scares the life out of me!"

"Would want assurances about controllers/ maintenance people being on hand at all times to fix any problems."

Q2 – Tasks to perform

1 – Substitution for tasks that humans currently perform, assistive tasks.

"Caring for the disabled, housework."

"Factories, Menial Jobs."

2 – Tasks that are difficult for humans to perform such as space exploration, or to hazardous for humans.

"Aid in more dangerous tasks such as firefighting, bomb disposal, or even representing a human in hostile conditions (humans being behind scenes)"

"Military"

Q2 – Tasks not to perform

1 – Tasks requiring humanity – caring, emotional support, decision making, education, medicine.

"Anything requiring empathy, compassion, thoughtfulness, discretion, lateral thinking, 'people skills'"

"Responsible jobs, decision making,"

2 – Other, including anti-social behavior and military robotics.

"Murder, crime"

"War, weapons"

Q3 – Sources

1 – Fictional Sources: TV-dramas, Films, Comics, Science Fiction Books

2 – Actual Source: Documentaries, Science Magazines, Academic Publications, News media

3 – Both Fictional and Actual Sources

4 – Non-classifiable.

The coding scheme was tested for reliability using inter-coder reliability measures. The results are presented in Table 1 and suggests that the scheme was reliable for this data.

Table 1. Coding Scheme

Questions		Code	Cohen's Kappa
Q1		**Positive**	**.859(36)**
	0	No Positive	
	1	Expectation of specific benefits, future possibilities	
	2	Other, including general positive sentiment	
Q1		**Negative**	**.734(44)**
	0	No negative sentiment	
	1	Negative changes to human mental and social processes (laziness, unemployment, meaning of humanity)	
	2	Others, including physical risks and maintenance costs	
Q2-Tasks		**Positive**	**.860(45)**
	1	Substitution for human tasks, including assistive roles	
	2	Others, including tasks that are hazardous for humans	
Q2 –Tasks		**Negative**	**.670(36)**
	1	Tasks requiring human-like qualities, decision making, caring, education, medicine.	
	2	Others, including anti-social, criminal and military tasks.	
Q3		**Sources**	**.816(39)**
	1	Fictional Sources, films, magazines, TV dramas	
	2	Real Sources, science books, TV news, actual experiences of robots	
	3	Both 1 and 2	
	4	Non-classifiable, eg just 'TV'	

3 Results

3.1 Demographics

The two samples were compared for systematic differences in terms of demographic characteristics. There were no significant differences between the two samples in terms of gender ($\chi^2(1)=.908, p>.34$). Examination of Table 2 suggests that there were no salient residuals, implying that the samples were equivalent in terms of gender.

Table 2. Gender Distribution

Sample	**Measure**	**Gender**		**Total**
UK		**Female**	**Male**	
	Count	77	53	130
	% Within Sample	59.2	40.8	100
	Std. Residual	.5	-.6	
Japan		**Male**	**Female**	
	Count	110	94	204
	%Within Sample	53.9	46.1	100
	Std. Residual	-.4	.4	
Total		**1**	**2**	
	Count	187	147	334
	%Within Sample	56.1	44	100

As for age, the mean age in the UK sample was 29.5 years while it was 24.1 in the Japanese sample. This difference was significant ($t(331)=4.77$, $p<.001$), and as such, one could not discount age as a systematic difference between the samples. Also, there were significant deviations from expected counts between the samples ($\chi^2(1)=7.07$, $p<.01$) in terms of number of students.. The salient residuals presented in Table 3 suggest that the UK sample had more non-students than the Japanese sample.

3.2 Independence Measures

Due to systematic differences between the samples in terms of age and student proportion, tests between the relationships between population characteristics and responses were run. The relationship between age and responses were assessed using a series of one-way ANOVAs, the results are presented in Table 4 . The results suggest that age did not have a systematic effect on the responses across the two samples.

A series of cross-tabulation analyses were performed in order to examine the impact of students within each sample. The results are presented in Table 5 and suggest that systematic variation between the samples would not be caused by the different proportions of students in the samples.

Table 3. Proportion of Students within the Samples

Sample	Measure			Total
UK		**Non-students**	**Students**	
	Count	53	77	130
	% Within Sample	40.8	59.2	100
	Std. Residual	1.7	-1.2	
Japan		**Non-students**	**Students**	
	Count	55	150	205
	% Within Sample	26.8	73.2	100
	Std. Residual	-1.4	.9	
Total		**1**	**2**	
	Count	108	227	335
	% Within Sample	32.2	67.8	100

Table 4. Age and Responses

Sample	Measure				
UK		**F**	**df**	**P**	**η^2**
	Positive Sentiment	.329	2(126)	.72	.005
	Negative Sentiment	.643	2(126)	.53	.010
	Task to perform	.801	1(120)	.37	.007
	Task not to perform	.299	1(104)	.59	.003
Japan		**F**	**df**	**P**	**η^2**
	Positive Sent	.076	2(197)	.93	.001
	Negative	.969	2(197)	.38	.010
	Task to perform	.208	1(184)	.65	.001
	Task not to perform	.098	1(96)	.76	.001

3.3 Cultural Comparisons

General Sentiments. There was significant deviation from expected counts between the samples ($\chi^2(2)$ =19.54,p<.001) in terms of positive sentiments. Examination of residuals in Table 6 suggest that participants in the UK referenced specific benefits (Code 1) to a lesser degree, and tended to reference more general feelings than the Japanese sample.

There was significant deviation from expected counts between the samples ($\chi^2(2)$=27.55,p<.001) for negative sentiments. Examination of the residuals in Table 7 suggest that participants from the UK sample were more likely to not reference negative sentiments (Code 0) than participants from the Japan sample at all. Also, participants from the UK Sample were less likely to reference emotional and social issues (Code 1) than participants from the Japanese Sample.

Table 5. Students vs non-students in responses

Sample	Measure				
UK		χ^2	**df**	**P**	**φ**
	Positive Sent	1.22	2	.54	.097
	Negative	1.41	2	.49	.104
	Task to perform	.084	1	.72	.026
	Task not to perform	2.59	1	.11	.156
Japan		χ^2	**df**	**P**	**φ**
	Positive Sent	1.61	2	.45	.089
	Negative	3.34	2	.19	.129
	Task to perform	.52	1	.47	.052
	Task not to perform	.15	1	.7	.039

Table 6. Positive Sentiments Towards Humanoid Robots according to Sample

Sample	Measure		Code		Total
UK		**0**	**1**	**2**	
	Count	60	35	35	130
	% Within Sample	46.2	26.9	22.8	100
	Std. Residual	.5	-2.2	2.6	
Japan		**0**	**1**	**2**	
	Count	83	95	23	201
	% Within Sample	41.3	47.3	11.4	100
	Std. Residual	-.4	1.8	-2.1	
Total		**0**	**1**	**2**	
	Count	143	130	58	331
	% Within Sample	43.2	39.3	17.5	100

Tasks Envisaged for the robot. There was no significant deviation between the samples in terms of what type of tasks the robot should perform ($\chi^2(1)=2.37,p>.1$). There were, however, significant deviations from expected counts between the samples($\chi^2(2)=18.56,p<.001$) in terms of tasks not to perform. Examination of the residuals in Table 8 suggest that participants from the UK referenced tasks requiring human-like qualities (Code 1) to a much larger extent than participants from the Japanese sample, while participants from the Japanese sample were more concerned with other aspects such as anti-social activities (Code 2).

Sources of information. There was significant deviation from expected counts between the samples ($\chi^2(3)=16.98,p<.001$). Examination of the residuals in Table 9 suggest that participants from the UK referenced getting information from both virtual and real sources (Code 3) to a much larger extent than participants from the Japanese sample.

Table 7. Negative Sentiments towards Humanoid Robots

Sample	Measure	Code			Total
UK		**0**	**1**	**2**	
	Count	67	17	46	130
	%Within Sample	51.5	13.1	35.4	100
	Std. Residual	2.5	-3.2	.3	
Japan		**0**	**1**	**2**	
	Count	59	76	66	201
	%Within Sample	29.4	37.8	32.8	100
	Std. Residual	-2.0	2.6	-.3	
Total		**0**	**1**	**2**	
	Count	126	93	112	331
	%Within Sample	38.1	28.1	33.8	100

Table 8. Activities not to perform

Sample	Measure	Code		Total
UK		**1**	**2**	
	Count	75	31	106
	%Within Sample	70.8	29.2	100
	Std. Residual	2.0	-2.2	
Japan		**1**	**2**	
	Count	40	58	98
	%Within Sample	40.8	59.2	100
	Std. Residual	-2.1	2.3	
Total		**1**	**2**	
	Count	115	89	331
	%Within Sample	56.4	43.6	100

Table 9. Sources of information

Sample	Measure	Code				Total
UK		**1**	**2**	**3**	**4**	
	Count	71	8	38	8	
	%Within Sample	56.8	6.4	30.4	6.4	100
	Std. Residual	-.9	-1.0	2.8	-.6	
Japan		**1**	**2**	**3**	**4**	
	Count	121	20	21	16	178
	% Within Sample	68	11.2	11.8	9	100
	Std. Residual	.8	.9	-2.3	.5	
Total		**1**	**2**	**3**	**4**	
	Count	192	28	59	24	303
	% Within Sample	63.4	9.2	19.5	7.9	100

3.4 Summary of Results

The general sentiment shows that participants from the UK sample were less likely to volunteer negative sentiment towards humanoid robots than in the Japanese sample. In terms of positive statements, the Japanese sample contained more statements regarding specific benefits that adoption of humanoids robots could provide. In terms of tasks that robots should not perform, the UK sample focuses on tasks that require "human-like" qualities, while the Japanese sample are more concerned like other things such as crime or war.

4 Discussion

The above results suggest that cultural differences do exist between a UK and a Japanese sample in terms of attitudes towards humanoid robots. Interestingly, these findings suggest that the Japanese sample was more likely to volunteer negative sentiments towards humanoid robots than the UK sample. The focus on emotional and social issues in the Japanese sample when expressing such negative sentiments are at odds with what one would expect from Kaplan's thesis on the Frankenstein Syndrome as a primarily Western concept.

This is in contrast to the results for tasks that participants would not want the robot to perform, where the UK sample's preferences suggested this sample either believed that humanoid robots could, or should, not perform tasks that required characteristics considered exclusive to humans by the sample; while the Japanese sample considered the role of humanoid robots as an extension of the general rules of society and warned against them being used in crimes, or violence.

5 Conclusions and Future Work

While there are clear cultural differences between the samples in terms of attitudes towards humanoid robots, these differences are complex and not clear-cut along a single dimension. We are currently preparing a new questionnaire, drawing on responses from both the UK and Japanese sample, which will allow for a rigorous quantitative exploration of these differences along several dimensions. The end result of this work will be a valuable tool for researchers investigating responses to robots across cultures.

Acknowledgements. This work was partly supported by EC integrated project LIREC (LIving with Robots and intEgrated Companions), funded by the European Commission under FP7-ICT, contract FP7-215554, and by the Japan Society for the Promotion of Science, Grants-in-Aid for Scientific Reseach No. 21118006. Dag Sverre Syrdal would like to thank Julie-ann Bell, Nikola C. Duncan and Paula Duncan for their help in distributing questionnaires.

References

[1] Kaplan, F.: Who is afraid of the humanoid? Investigating cultural differences in the acceptance of robots. International Journal of Humanoid Robotics, 1–16 (2004)
[2] Shelley, M.: Frankenstein, or, The modern Prometheus, Rev. ed. Penguin Books, London (2003)
[3] Scott, R.: Blade Runner. Warner Bros (1982)
[4] Stanton, A.: Wall-E. Pixar-Disney (2008)
[5] Mori, M.: The Buddha in the robot, 1st edn. Kosei Pub. Co., Tokyo (1981)
[6] Bartneck, C., Nomura, T., Kanda, T., Suzuki, T., Kensuke, K.: Cultural differences in attitudes towards robots. In: Proceedings of the AISB Convention: Symposium on Robot Companions: Hard Problems and Open Challenges in Human-robot Interaction. University of Hertfordshire, Hatfield (2005)
[7] Bartneck, C., Nomura, T., Kanda, T., Suzuki, T., Kennsuke, K.: A cross-cultural study on attitudes towards robots. In: Proceedings of the HCI International 2005, Las Vegas (2005)
[8] Bartneck, C., Suzuki, T., Kanda, T., Nomura, T.: The influence of people's culture and prior experiences with Aibo on their attitude towards robots. AI & Society 21(1-2), 217–230 (2006)
[9] MacDorman, K.F., Vasudevan, S.K., Ho, C.-C.: Does Japan really have robot mania? Compar-ing attitudes by implicit and explicit measures. AI & Society 23(4), 485–510 (2008)
[10] Nomura, T., Kanda, T., Suzuki, T., Kato, K.: Psychology in Human-Robot Communication: An Attempt through Investigation of Negative Attitudes and Anxiety toward Robots. In: Proc. 2004 IEEE International Workshop on Robot and Human Interactive Interactive Communication, Kurashiki, Okayama, Japan, September 20-22, pp. 35–40 (2004)
[11] Allport, G.W.: Traits revisited. American Psychologist 21(1), 1–10 (1966)
[12] Syrdal, D.S., Dautenhahn, K., Koay, K.L., Walters, M.L.: The negative attitudes towards robots scale andreactions to robot behaviour in a live human-robotinteraction study. In: Proc. of the AISB Symposium on New Frontiers in Human-Robot Interaction (2009)

The Effects of a Robot Instructor's Positive vs. Negative Feedbacks on Attraction and Acceptance towards the Robot in Classroom

Eunil Park[1], Ki Joon Kim[1], and Angel P. del Pobil[1,2]

[1] Department of Interaction Science, Sungkyunkwan University, Seoul, South Korea
{pa1324,veritate}@skku.edu
[2] Robotic Intelligence Laboratory, University Jaume-I, Castelló, Spain
pobil@icc.uji.es

Abstract. As robots are now being widely used as educational aids and assistants, it is crucial to understand the effects of robotic teaching assistants in classroom and how attraction and acceptance towards the robot are shaped. A 2 (type of instructor: human vs. robot) x 3 (feedback style: positive vs. negative vs. neutral) between-subjects experiment with six conditions was conducted to examine the effects of a robot instructor in classroom and the instructor's feedbacks on students' attraction and acceptance towards the given feedback. Results showed that feedback from a human instructor were more acceptable than feedback from a robot instructor. Students in the robot-instructor condition showed greater attraction towards the instructor when received a positive feedback, whereas students in the human-instructor condition did not report any difference in their attraction towards the instructor due to the feedback style. Both implications and limitations of the present study as well as guidelines for future research are discussed.

Keywords: robot teacher, robot instructor, HRI, education, feedback.

1 Introduction

The purpose of the present study was to examine whether attraction and acceptance towards a robot instructor are influenced by type of feedback (positive, negative, or neutral) given by the robot instructor in a classroom environment. Findings from previous studies suggest that robots play various and important roles in our life [1]. It is therefore essential to further explore interactions between humans and robots in diverse contexts (i.e., classroom) [2], [3], [4], [5] and [6].

1.1 Robot as Instructor

Improved efficiency of robots and rapid advancement of robotic technology in recent years have allowed robots to be used for various purposes, including education. Unlike human instructors, robots never get tired [7]. Some studies attempted to use

B. Mutlu et al. (Eds.): ICSR 2011, LNAI 7072, pp. 135–141, 2011.

robots in educational environments. For instance, specially designed robots provide the concept of R-learning in elementary school, kindergarten or home [8]. Although robot instructors can provide students with more professional, limitless knowledge, robot instructors are still unable to fully replace human instructors because robots have limited capability to self-consciously deliver ideal feedbacks at right times [9]. Students are influenced not only by verbal feedback but also by non-verbal feedback such as instructors' facial expressions and gestures, which only human instructors are able to adequately provide. Therefore, this study was conducted to explore by following a research question:

Research Question 1. What is the relationship between instructor type (Independent Variable: human vs. robot) and attraction and attitudes towards the instructors (Dependent Variable)?

1.2 Positive, Negative, or Neutral Feedback

Studies based on the Computer as Social Actor (CASA) paradigm suggest that people are more likely to assess flattering computers more positively. People perceive computers giving positive feedbacks to be more socially attractive, while perceiving computers with negative feedbacks to be less attractive. If the CASA paradigm is applicable to HRI, then robots that give positive, encouraging feedbacks may be perceived as more attractive than robots that give negative feedbacks [10], [11]. The present study thus intends to explore whether a robot's demeanor follows social rules of human-human interaction by examining the following research question:

Research Question 2. What is the relationship between feedback style (Independent Variable: positive vs. negative vs. neutral) and level of acceptance towards the given feedback (Dependent Variable)?

2 Method

The experiment was between-subjects design with six conditions: 2 (type of instructor: human vs. robot) x 3 (feedback style: positive vs. negative vs. neutral).

2.1 Participants

126 students were recruited from a university in Seoul. The age of the participants ranged from 19 to 33 (M=23.48, SD=2.60). Each condition had 21 participants with 10 males.

2.2 Apparatus and Stimulus Material

In the human-instructor condition, an experimenter prepared to teach with a desktop computer connected to a vim projector. A graduate student majoring in history was trained for a week as a human instructor for the experiment. In the robot-instructor condition, Nao, a humanoid robot [12] (Fig. 1), was seated at a desk and posed as an instructor.

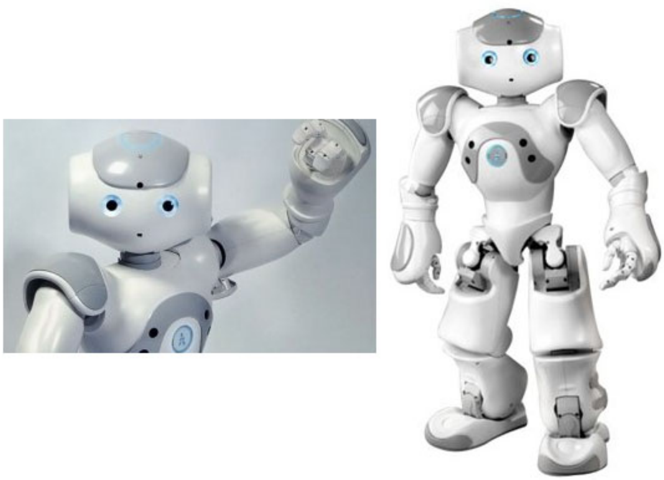


Fig. 1. Nao Robot, used in our experiment

Twenty respondents participated in a pretest to select stimulus material which was neutral and insensitive in order to avoid potential story-specific and prejudiced effects. We initially chose six articles from introductory-level college textbooks. Respondents were asked to read the articles and then complete a questionnaire survey on 7-point Likert scales measuring the neutralness of each article. Based on the results of questionnaire survey, we chose an article that was rated as having the most neutral content among the six articles (4.2 on a 7-point scale). The selected article contained a history topic about the Renaissance era [9] and [13]. It was a cultural movement in Europe in 14^{th} century started from Italy, which affected various disciplines such as literature, philosophy, music, etc.

In order to eliminate the effect of robot's additional characteristics such as gesture and voice, we used an audio file decoded by the human-instructor. That is, the voices of human instructor and robot instructor were identical. The robot's gestures were adopted from the human instructor. The humanoid robot has 25 degree of freedom. Software called NaoQi was used to program and control Nao's sensors and motors. The NaoQi API provides a simple interface programmable with C++ and Python. This API allows to control the robot's actuators, movements, and speech function.

2.3 Procedure

To minimize the effect of pre-existing knowledge of the contents, we excluded history majors from our participant pool.

Upon arrival at the laboratory, participants were told that they were going to attend a class for about 10 minutes. Students were seated in a way that they were able to see and listen to the instructor as well as the learning contents being projected via a vim projector.

After the session was over, participants were asked to answer a short quiz (posttest) about the contents they just learned. The instructor then gave participants either positive, negative, or neutral feedback on the quiz score, regardless of participants' actual scores.

2.4 Measurement

Social attraction towards the instructors (Cronbach's α=.77) was an index composed of eight items previously used by [10] and [14]. An index composed of four items measuring the level of acceptance towards the instructor's feedback (on, α=.84) was administered in order to evaluate participants' attitudes towards the received feedback, [10], [14]. Participants responded to each item by marking on a 7-point Likert scale, ranging from 1="strongly disagree," to 7="strongly agree."

Table 1. Dependent variables and questionnaire items

Questionnaire Category	Description
Social attraction towards the instructor	How much did you like the instructor? How much did you like interacting with the instructor? How much did you think the instructor was cheerful? How much did you think the instructor was cooperative? How much did you think the instructor was friendly? How much did you think the instructor was happy? How much did you think the instructor was kind?
Acceptance towards the instructor's feedback	The instructor's feedback was acceptable. The instructor's feedback was accurate. I liked being evaluated by the instructor. I will let my friends know about the feedback I received from the instructor.

2.5 Results

A series of 3 x 2 factorial analyses of variance (ANOVA) was conducted to analyze the effects of feedback style and instructor type on the dependent variables, followed by post-hoc analyses using Student's t test. The results from the ANOVA and a subsequent post-hoc analysis indicated that participants taught by the human instructor (M=5.01, SD=1.12) reported a significantly higher degree of social attraction towards the instructor than those in the robot-instructor condition (*M*=4.16, *SD*=1.59), $F(1, 120)=17.066$, $p<.001$. The instructor type also had a significant effect on the level of acceptance towards the instructor's feedback. Participants reported that the feedback from the human instructor (*M*=4.89, *SD*=1.18) was more acceptable than the feedback from the robot instructor (*M*=4.17, *SD*=1.34), $F(1, 120)=14.183$, $p<.001$.

Feedback type also had effects on social attraction and the level of acceptance towards the instructor's feedback; participants who received a positive feedback (*M*=5.46, *SD*=1.17) and neutral feedback (M=4.63, SD=1.37) reported a significantly higher degree of social attraction than those received a negative feedback did (*M*=3.67, *SD*=1.20), $F(1, 120)=25.167$, $p<.001$. In addition, participants who received a positive feedback (*M*=5.20, *SD*=1.10) and neutral feedback (*M*=4.79, *SD*=1.13) reported a significantly higher level of acceptance towards the instructor's feedback than those who received a negative feedback (*M*=3.67, *SD*=1.16), $F(1,120)=25.071$, $p<.001$.

Interaction between feedback style and instructor type on social attraction was significant, $F(2, 120)=3.206$, $p=.044$ (Fig. 2 and 3), such that participants in the robot-instructor condition showed greater social attraction towards the instructor when received a positive feedback. Participants in the human-instructor condition, on the other hand, did not report any difference in their attraction towards the instructor due to the feedback style.

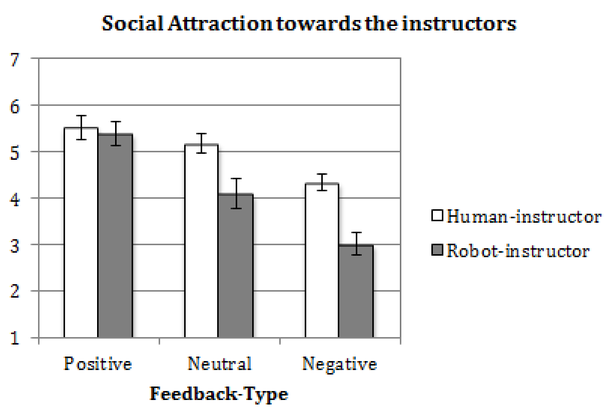


Fig. 2. Mean of social attraction towards the instructors

3 Discussion and Conclusion

Our findings convey valuable insights for both robots in educational environments and future work pertaining to students' acceptance and attitudes towards robot instructors in classroom. In our previous work investigating a robotic teacher [13], we witnessed a robot's full potential of being an effective instructor, and the present study further demonstrated a possible application of robots to education by examining students' perceived attraction to and attitudes towards the robot instructor. Although robot instructors are still unfamiliar to most students, they have much potential for next generation of commercial education [9] by reducing human resources and introducing diverse teaching methods.

There were some limitations in our study. In our experiment, the robot instructor was relatively expressionless compared to the human instructor. The robot instructor was not able to provide as much facial expressions and gestures as the human instructor did, ignoring the importance of non-verbal communication. If our experiment was conducted with a robot with ability to instantly response and make facial expressions, we might have found different results. Future studies may be further strengthened by examining the role of attitudes towards technology acceptance and other classroom settings such as kindergarten and high school.

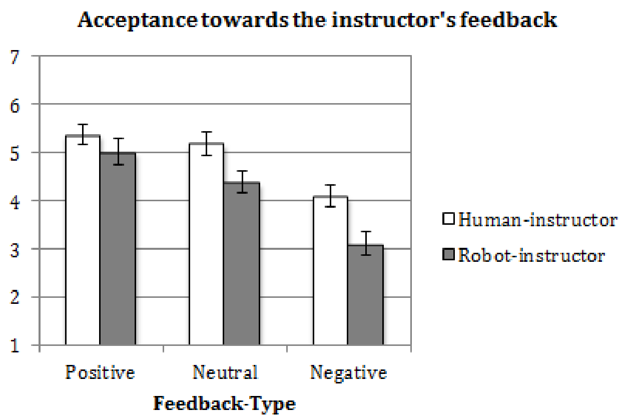


Fig. 3. Mean of social attraction towards the instructors

Acknowledgments. This study was supported by a grant from the World-Class University program (R31-2008-000-10062-0) of the Korean Ministry of Education, Science and Technology via the National Research Foundation.

References

1. Fong, T., Nourbakhsh, I., Dautenhahn, K.: A survey of socially interactive robots. Robotics and Autonomous Systems 42(3), 143–166 (2003)
2. Siegel, M., Breazeal, C., Norton, M.I.: Persuasive robotics: the influence of robot gender on human behavior. In: IEEE/RSJ International Conference on Intelligent Robots and Systems 2009, pp. 2563–2568. IEEE Press, New York (2009)
3. Kim, H., Bae, K., Yoon, H.: Age and gender classfication for a home-robot service. In: 16th IEEE International Symposium on Robot and Human Interactive Communication 2007, pp. 122–126. IEEE Press, New York (2007)

4. Dautenhahn, K., Walters, M., Woods, S., Koay, K.L., Nehaniv, C.L., Sisbot, A., Alami, R., Simeon, T.: How I serve you?: a robot companion approaching a seated person in helping context. In: 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction, pp. 172–179. ACM, New York (2006)
5. Walters, M.L., Syrdal, D.S., Koay, K.L., Dautenhahn, K., te Boekhorst, R.: Human approach distances to a mechanical-looking robot with different robot voice styles. In: 17th IEEE International Symposium on Robot and Human Interactive Communication 2008, pp. 707–712. IEEE, New York (2008)
6. Walters, M.L., Dautenhahn, K., te Boekhorst, R., Koay, K.L., Woods, S.N.: Exploring the Design Space of Robot Appearance and Behavior in an Attention-Seeking 'Living Room' Scenario for a Robot Companion. In: 2007 IEEE Symposium on Artificial Life, pp. 341–347. IEEE, New York (2007)
7. Goodrich, M.A., Schultz, A.C.: Human-robot interaction: a survey. Foundations and Trends in Human-Computer Interaction 1(3), 203–275 (2007)
8. Han, J., Jo, M., Park, S., Kim, S.: The Educational Use of Home Robots for Children. In: 14th IEEE International Workshop on Robot and Human Interactive Communication 2005, pp. 378–383. IEEE, New York (2005)
9. Park, E., Kim, K.J., del Pobil, A.P.: Is teaching-robot possible?: Practical Study for teaching-robot. In: 16th IASTED International Conference on Robotics and Applications (IASTED- RA 2011), pp. 132–137. ACTA Press, Calgary (2011)
10. You, S., Nie, J., Suh, K., Sundar, S.S.: When the robot criticizes you...Self-serving bias in human-robot interaction. In: 6th International Conference on Human-Robot Interaction, pp. 295–296. ACM, New York (2011)
11. Fogg, B., Nass, C.: Silicon Sycophants: the Effects of Computers that flatter. International Journal of Human-Computer Studies 46(5), 551–561 (1997)
12. Aldeberan, Nao Robot, http://aldeberan.com
13. Park, E., Kong, H., Lim, H., Lee, J., You, S., del Pobil, A.P.: The Effect of Robot's Behavior vs. Appearance on Communication with Humans. In: 6th ACM/IEEE International Conference on Human Robot Interaction (HRI 2011), pp. 219–220. ACM, New York (2011)
14. Groom, V., Chen, J., Johnson, T., Kara, F., Nass, C.: Critic, compatriot, or chump?: Responses to robot blame attribution. In: 5th ACM/IEEE International Conference on Human Robot Interaction (HRI 2010), pp. 211–218. ACM, New York (2010)

Engkey: Tele-education Robot

Sangseok Yun[1], Jongju Shin[2], Daijin Kim[2],
Chang Gu Kim[1], Munsang Kim[1], and Mun-Taek Choi[1]

[1]Center for Intelligent Robitics at Korea Institute Science and Technology,
Hwarangno 14-gil 5, Seongbuk-gu, Seoul 136-791, Korea
{ssyun,cgkim93,munsang,mtchoi}@kist.re.kr
[2]Department of Computer Science and Engineering, POSTECH,
San 31, Hyoja-Dong, Nam-Gu, Pohang 790-784, Korea
{jjshin,dkim}@postech.ac.kr

Abstract. In this paper, we introduce a new form of an English education system that utilizes a teleoperated robot controlled by a native human teacher in a remote site. By providing a unique operation interface that incorporates non-contact vision recognition technologies, a native teacher easily and naturally controls the robot from a distance while providing lectures. Due to its effective representation of a human teacher in a robot, students show great interest on the robot and feel comfortable learning English with the teleoperated robot. In a real field pilot study, the participated elementary students have achieved good improvements on standardized tests after the study, which shows the effectiveness of the teleoperated robot system.

Keywords: Educational robot, Engkey, Distance education, Avatar assistant robot.

1 Introduction

Service robots are highly applicable to assisting humans in the areas of education, elderly care, guidance, home service, and so forth [1] [2] [3]. In addition, a variety of applications have been explored in teleoperated robotics [4]. In many cases, the applications require dedicated hardware for special purposes and highly trained operators to control and interact with the mechanisms in the remote environment. In other case, Personal Roving Presence (PRoP) provides simple, inexpensive, and natural human communication and interaction [5]. The virtual talking heads concern the use of advanced natural interface with emotional speech and facial gesture in tele-education [6].

With a distant-teaching concept, it seems feasible to suggest a new paradigm to the current English (or any other languages) education system by providing a teleoperated robot solution, controlled remotely by a native teacher in an English-speaking country. The main intention is that students at schools who have little experience with native teachers for economic and other reasons will benefit from

B. Mutlu et al. (Eds.): ICSR 2011, LNAI 7072, pp. 142–152, 2011.

this type of direct contact with a native instructor through a robot. Although current robotic technologies are not mature enough to suppress human teachers from classrooms, it seems possible to provide a cost-effective way of education.

To achieve this goal, it is a key element that the teleoperated system should provide a natural unity between a native teacher and robot, so that students can concentrate and be fully immersed in the learning process. We are attempting to achieve this unity by developing an easy and real-time control of the robot's various activities and expressions by a native teacher as if the robot were the teacher's avatar.

To guarantee such a unity, significant integration of various robotic technologies is essential. Transmitting teacher and student's video and audio data bilaterally between remote to local space without interruption is crucial for the quality of the lecture. Safety is one of the most critical issues for robots especially when the robot dynamically interacts with little kids. Since the teacher in a remote site commands the robot based on very limited information on how the robot actually moves or behaves, the robot should have full safety features such as autonomous navigation in the classroom with obstacle avoidance and emergency stop. A user interface (UI) of the teacher in a remote site should be fairly easy and natural to convey teacher's expression properly. It seems to be a natural choice to utilize vision technologies such as facial expression and head-hand gesture recognition for the interface so that the teacher does not need to wear additional equipment. The teacher's natural movement during teaching seamlessly transferred to robot's movement.

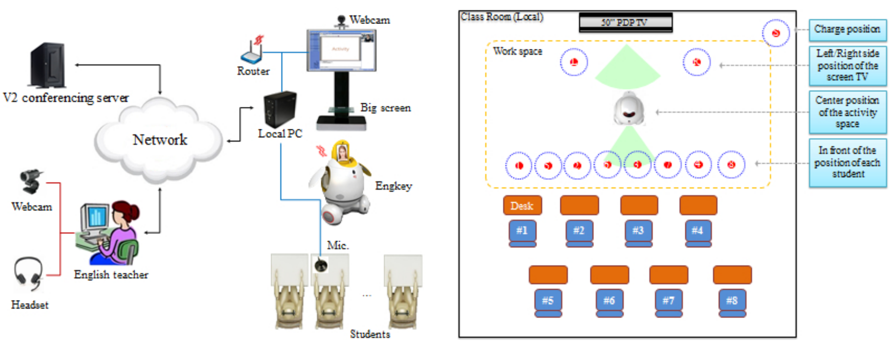
The second goal of this paper is to develop an educationally acceptable robot character by utilizing interest and curiosity about the robot. From the class, the robot provides educational supplement in an interesting way through conversation and practice such as pronunciation practice, sing-alongs and robotic games. In particular, robotic contents including dynamic behavior on pan-tilt and arm, various emotional expression on its avatar lip syncing and LED, and free movement with reactive safety and omni-directional wheel easily make students more focused and memorize an English song or rhyme and sing along.

This paper is organized as follows. Section 2 describes the description of telepresence-type English teaching robot system, Section 3 illustrates user interface related with remote robot control, Section 4 presents the details of facial imitation, Section 5 introduces robot navigation in classroom, Section 6 demonstrates experimental result, and Section 7 concludes the paper.

2 System Configuration

As a distance-education environment, the main role for teaching methodologies is that remote native teacher should focus on education contents with a minimal operational intervention and a serviceable control tool as well as real-time avatar animation system executes teacher's intention in the local place. While remote-side video and audio stream links to the local space, remote-side teacher interacts with local students by using an interactive tool. Especially, remote user interface is able to control the class including song/chant and robot functionality such as pre-defined robotic expression and designated robot position. On the other hand, teacher put facial expression, hand gesture, and speaking to the students by mapping between remote visual/vocal recognition and local facial imitation/lip-syncing.

For the local site, the classroom consists of 8 students seats (actually it can be more), traversable robot space, and big screen. As a local teaching assistant, the robot proceeds with lesson, navigates to target position promptly where the teacher wants in pre-built map by omni-directional robot wheel, and interacts with students including human tracking. Big screen display the teaching material. The external microphone on the desk gathers the student's voice. And two cameras on the big screen and attached robot head get the viewing for classroom configuration and students location. Figure 1 shows the telepresence-type English teaching robot system.



(a) System configuration. **(b)** Local classroom.

Fig. 1. Telepresence-type English teaching robot system

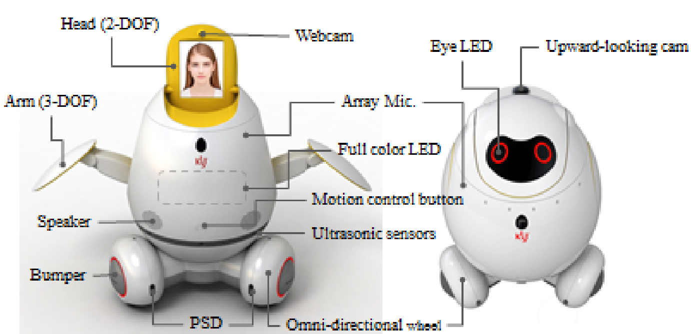


Fig. 2. System overview of Engkey

In the classroom, a certain distance away from the big screen, where students are located on. And the robot to move between the big screen and students will have the space. we set 12 specified destination in map building process. 8 positions are the location of students and the rest are center position, TV front left and right, and charge position as shown in Fig. 1b.

Engkey (namely English jockey), developed in the Center for Intelligent Robotics (CIR), provides educational assistance at elementary school, remote interactive ability, and mental/emotional support such as being a teacher for conversation and entertainment. Consequently, Engkey is mainly organized as follows:

1. Avatar assistant robot, which can be controlled remotely by a native teacher, animates the 3D face model with facial expression and lib-sync for remote user's voice.
2. Engkey has a 2-DOF head, two 3-DOF arms in order to show various robot emotion expressions on its head monitor and LED.
3. Engkey traces a local user's call where the sound is generated and keeps an eye-contact by using human detector and pan-tilt unit automatically.
4. Engkey steers clear of the obstacles by measuring the distance through sonar sensor and move to the user's desired goal by using localization based on ceiling map.
5. Engkey has reactive behavior with safety. If the bumper is pressed, the robot stop and fold the head and arm in order to avoid unexpected human touch.
6. Engkey provides immerse robot education continuously by robot design as a close friend and free motion with omni-directional wheel.

Figure 2 shows main functionalities and specifications of Engkey.

3 Remote User Interface

One of the main purposes in the distance education system is to make real-time character animation system requiring a minimal intervention of human animators. We propose remote user interface with semi-autonomous robot behavior so as to implement remote teacher's intend as shown in Fig. 3. this interface consists of three different functionalities to communicate with each other.

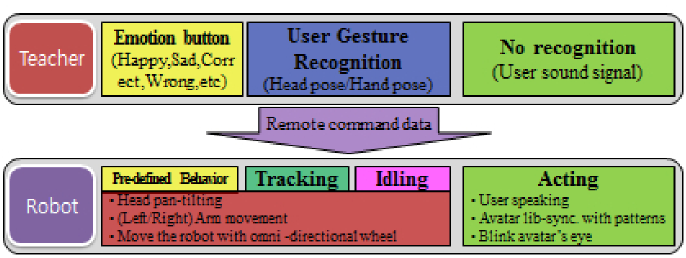


Fig. 3. Semi-autonomous robot behavior by remote teacher's intention

First, remote teacher directly control the robot in local classroom by using robot control panel. It is composed of six control categories. In lesson, Teacher is able to handle remote connectivity, lesson start/stop, and selection for song/chant easily, shows emotion expression as a reaction for student's activity, and moves to appointed position. Figure 4 shows remote robot control panel with six control categories.

Second, a remote user interface including user gesture recognition in remote system needs a non-contact human data. And, this acquires teacher's face image, recognized head/hand position, and facial features from the image plane as well as recognized data and teacher's voice transmit to the local system simultaneously. These command data make the robot behavior with unconditional acting according to the state. In last, network video-conferencing system allows to access the conference to teach English and share the educational contents with local students. Remote teacher safely receives local student's visual/vocal stream over the internet in real time.

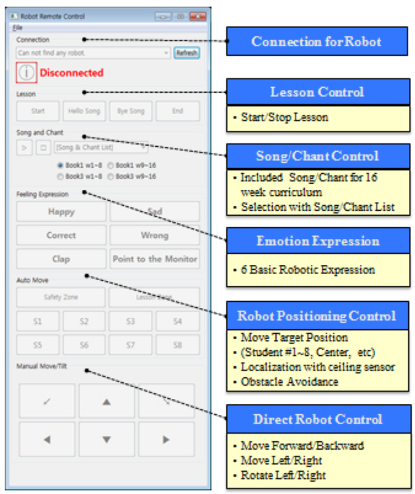


Fig. 4. Remote robot control panel

4 Face Imitation

In this work, we use two algorithms for facial imitation: the first is head tracking algorithm, and the second is facial feature tracking algorithm. Head tracking algorithm is used to imitate user's head pose, and facial feature tracking algorithm is used to imitate user's eye and mouth's movements.

4.1 Head Tracking

In this section, we use 3D cylindrical head model which is introduced by [7] [8] [9]. Especially, Ryu [9] proposed illumination robust head tracking algorithm by adding illumination basis vectors to the template. This approach can cover the rapidly changing illumination. The objective is then to find the parameters which minimize the function given by

$$\sum_x \left[I(W(X;p),t) - \sum_i^{bN} (q_i + \triangle q_i) b_i(X) - I(W(X;p+\triangle p), t+1) \right], \quad (1)$$

where p and q_i are the 3D rigid motion parameter and the i-th illumination coefficient, respectively, and $\triangle p$ and $\triangle q_i$ are the updated parameters computed by solving optimization problem.

Fig. 5. Definition of facial features: 4 for eye brows, 8 for eyes, and 10 for mouth

4.2 Facial Feature Tracking

For the facial feature tracking, we represent the face with 22 feature points as shown in Fig. 5. The results for the AAM fitting designate the initial positions of the 22 feature points. And we define an 11x11 pixel tracking block centered at each feature point. The 22 feature points are tracked by the model-based Lucas–Kanade feature tracker[10], which is introduced by Shin et al. [11]. In [11], the proposed algorithm is focused on the lip feature tracking. We extend the algorithm to the facial feature tracking. To represent the face, we use shape model like as

$$W(x;p) = s_0(x) + \sum_{i=1}^{n} p_i s_i(x), \quad (2)$$

where x is the position on the mean shape, $x_i(x)$ is the i-th basis. Then the face is represented by the linear combination of the selected bases. A more complete face shape model may be found in [11].

5 Navigation

Autonomous navigation in the classroom mainly considers two parts of rapid localization and dynamic obstacle avoidance based on the map. We propose a novel upward-looking camera-based localization scheme using corner, lamp features [12], and fiducial [13]. The originality of this scheme lies in the detection method for ceiling landmark, and the method for combining this feature information with accumulated robot movement in the localization process to achieve stable navigation. So, we used an extended Kalman filter (EKF). The state vector is defined as

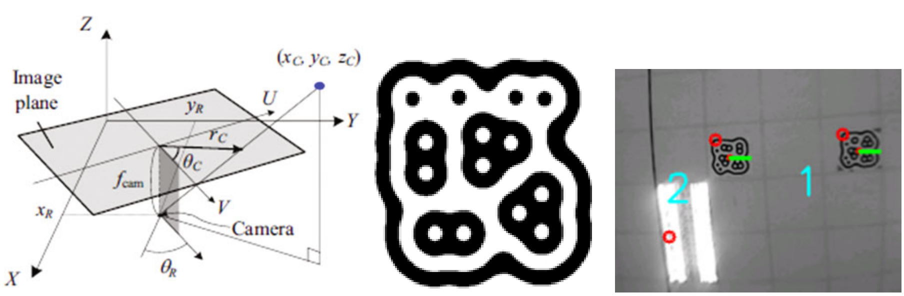
$$X = \left[X_R, X_{F_i}, X_{C_j}, X_{L_k}\right]^T, \tag{3}$$

where

$$\begin{array}{ll} X_R = [x_R, y_R, \theta_R], & X_{F_i} = [x_{F_i}, y_{F_i}, \theta_{F_i}], \\ X_{C_j} = \left[x_{C_j}, y_{C_j}, \theta_{C_j}\right], & X_{L_k} = [x_{L_k}, y_{L_k}, \theta_{L_k}], \end{array} \tag{4}$$

where X_R is the robot pose, and X_{F_i}, X_{C_j}, and X_{L_k} are the positions of fiducial i, corner j, and lamp k, respectively. The projected points for fiducial, corner, and lamp were denoted as 2-D points (r_F, θ_F), (r_C, θ_C), and (r_L, θ_L) in the image plane. The procedure of the EKF consists of the prediction and update steps in order to estimate the robot pose and ceiling feature positions. For the first state or kidnapped situation, rapidly global localization is a very crucial point. Proposed approach tries to find the ceiling landmark. Especially recognizing fiducial marker in the image plane directly gets the position where it is. Fig. 6a shows the observation models. Fig. 6b shows one of the fiducial markers. And, Fig. 6c shows the detected ceiling landmarks. A more complete ceiling feature concept and recursive techniques may be found in [12].

In order to set the target position and environment modeling directly in various classroom environments, we provide a map building process by using upward-looking camera and sonar sensors for all directions. Particularly standardized modeling classroom environment makes it easy to build a map. The component of map builder build an occupancy map using sonar sensors, enroll the visual feature such as corner, lamp, and attached fiducials on the ceiling automatically according to the robot movement. User just moves the robot to unorganized area and sets the goal position by using user interface. Figure 8 shows the Graphical User Interface for map building with ceiling fiducials' position (blue), corner and lamp position (red), and enrolled goal position (green).



(a) Observation models. **(b)** Fiducial marker. **(c)** Ceiling landmarks.

Fig. 6. Navigation with ceiling landmarks

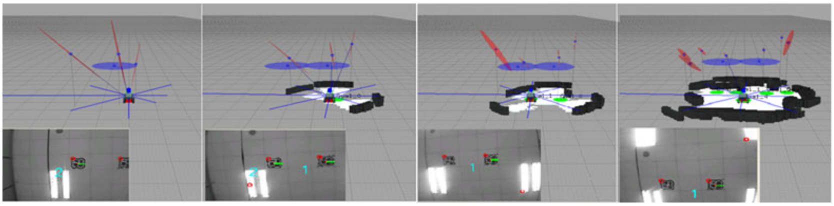


Fig. 7. The Graphical User Interface for map building

6 Experimental Result

A variety of components for the robot performance was tested in a distant teaching robot system. Firstly, as a substantial representation of the user's emotions, we have five configurations of the emotional expression such as neutral, correct, sad, happy, wrong state. User in two way can express the feelings. One is clicking the emotion expression button in the remote robot control panel. Another is showing user's faces on the camera to track facial feature. In the latter case, it needs to initialize the feature model and fit face for the camera within 2 seconds. Once feature model fits into the face, user can move the head and facial feature points track user face within 15 degree for any direction. Table 1 illustrates the tracking performance for the facial feature points and head model. For tracking state, the resolution of the head feature points has within 1cm for positions and 5 degree for angles. And, each of five facial expressions can be seen that well-implemented as shown in Fig. 8a.

Table 1. Tracking performance for facial feature and head

	Initializing	Tracking
Min. face size (*pixels*)	50	50
Max. face occlusion (*pixels*)	0	10
Head yawing (*deg.*)	< 15	$-80 < X < 80$
Head pitching (*deg.*)	< 15	$-30 < Y < 60$
Head rolling (*deg.*)	< 15	$-80 < Z < 80$
Facial feature yawing (*deg.*)	< 15	$-15 < X < 15$
Facial feature pitching (*deg.*)	< 15	$-15 < Y < 15$
Facial feature rolling (*deg.*)	< 15	$-15 < Z < 15$
Time to acquire (*sec*)	1–2	.
Position error (*cm*)	.	1
Orientation error (*deg.*)	.	5

As a second experiment, we tested a hand/head pose detection and mapping to the robot for detected data. Especially 3D cylindrical model generates pitch-yaw angle for robot head in 2D image coordinates at the rate of 3rad/s. Likewise, detected hand position match up with robot arm by considering 3D kinematics. Basic activity of the robot through user's behavior imitation was that it was

(a) The avatar representation of facial feature tracking

(b) A detected human data and match up with the robot

Fig. 8. Robotic expressions through user's behavior

possible. Figure 8b shows the detected human feature for natural behavior and match up with the robot.

A pilot study testing the tele-education robot, Engkey, in real fields has been performed for three months in 2011. A total of 29 robots have been deployed in the 29 English classrooms of the elementary schools in the City of Daegu, Korea. Each native teacher in the Philippines conduct classes and 8 students in each class participated in the class. Figure 9 shows the actual scene for tele-education class. Fig 9a shows classroom configuration with avatar assistant robot and big screen. Fig 9b shows the robot interacting with students. And fig 9c shows native teacher operating a class material and a remote robot control panel.

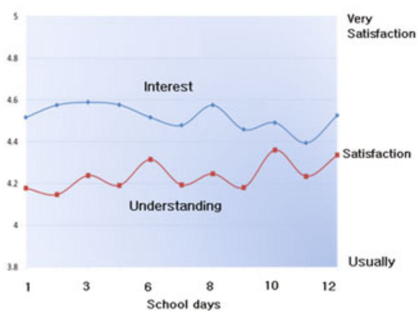
(a) Robot and big screen

(b) Interacting with student

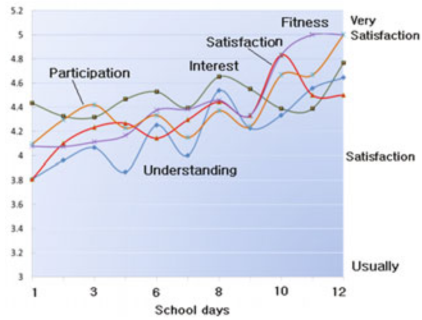
(c) Remote lesson progress by teacher

Fig. 9. A scene of tele-education

In the pilot study, the teacher in a remote site actually teaches English to students in the 12-week course including word robotic song-and-chants and games. As can be seen in Fig. 10, the assessment for English lessons before and after shows the student's achievement and teacher feedback for participation, interest, satisfaction, fitness, and understanding. During three months, students have improved English language skill as well as an interest in robot. Teacher feedback also shows that five assessment items for students gradually have increased. As a result, the tele-education robot was found to be more effective and interesting to students during the pilot project.



(a) Students feedback.



(b) Teacher feedback.

Fig. 10. Interest and satisfaction analysis of a pilot study

7 Conclusion

In this paper, we introduce a new form of an English education system that utilizes a teleoperated robot controlled by a native human teacher in a remote site. By providing a unique operation interface that incorporates non-contact vision recognition technologies, a native teacher easily and naturally controls the robot from a distance while providing lectures. The robot is equipped with the navigation with collision avoidance for maximum safety, human-friendly expression behaviors and so forth. Due to its effective representation of a human teacher in a robot, students show great interest on the robot and feel comfortable learning English with the teleoperated robot. In the real field tests with 690 elementary students for 3 months in Korea, students have achieved good improvements especially on speaking according to teacher's feedback, which shows the effectiveness and possibility of the system. Although the robotic tele-education did not intend to match or surpass the performance of direct human teaching, it demonstrated that it could be a cost-effective way of education.

Acknowledgments. This research was performed for the Intelligent Robotics Development Program, one of the 21st Century Frontier R&D Programs funded by the Ministry of Knowledge and Economy of Republic of Korea.

References

1. Thrun, S., Bennewitz, M., Burgard, W., Cremers, A.B., Dellaert, F., Fox, D., Hahnel, D., Rosenberg, C., Roy, N., Schulte, J., Schulz, D.: MINERVA: A Second-Generation Museum Tour-Guide Robot. In: International Conference on Robotics and Automation, pp. 10–15 (1999)
2. Hans, M., Baum, W.: Concept of a Hybrid Architecture for Care-O-bot. In: International Workshop on Robot-Human Interactive Communication, pp. 407–411 (2001)

3. Yun, S., Kim, C.-G., Choi, M.-T., Kim, M.: Robust Robot's Attention for Human Based on the Multi-modal Sensor and Robot Behavior. In: IEEE Workshop on Advanced Robotics and its Social Impacts, pp. 117–122 (2010)
4. Sheridan, T.B.: Telerobotics, Automation, and Human Supervisory Control. MIT Press (1992)
5. Paulos, E., Canny, J.: Social Tele-Embodiment: Understanding Presence. Autonomous Robots 11, 87–95 (2001)
6. Bonamico, C., Lavagetto, F.: Virtual Talking Heads for Tele-education Application. In: SSGRR (2001)
7. Cascia, M., Sclaroff, S., Athitsos, V.: Fast, reliable head tracking under varying illumination: An approach based on robust registration of texture-mapped 3d models. IEEE Trans. Pa. Anal. Mach. Intel. 22, 322–336 (2000)
8. Xiao, J., Moriyama, T., Kanade, T., Cohn, J.: Robust full-motion recovery of head by dynamic templates and re-registration techniques. Int. J. Imag. Syst. Tech. 13, 85–94 (2003)
9. Ryu, W., Kim, D.: Real-Time 3D Head Tracking Under Rapidly Changing Pose, Head Movement and Illumination. In: Kamel, M.S., Campilho, A. (eds.) ICIAR 2007. LNCS, vol. 4633, pp. 569–580. Springer, Heidelberg (2007)
10. Baker, S., Matthews, I.: Lucas-Kanade 20 Years On: A Unifying Framework. Int. J. Comp. Vision. 56(3), 221–255 (2004)
11. Shin, J., Lee, J., Kim, D.: Real-Time lip reading system for isolated Korean word recognition. Pattern Recognition 44(3), 559–571 (2011)
12. Hwang, S.-Y., Park, J.-T., Song, J.-B.: Autonomous Navigation of a Mobile Robot Using an Upward-Looking Camera and Sonar Sensors. In: IEEE Workshop on Advanced Robotics and its Social Impacts, pp. 40–45 (2010)
13. Costanza, E., Robinson, J.: A Region Adjacency Tree Approach to the Detection and Design of Fiducials. In: Vision, Video and Graphics, pp. 63–70 (2003)

Motivating Children to Learn Arithmetic with an Adaptive Robot Game

Joris B. Janssen[1], Chrissy C. van der Wal[1,2],
Mark A. Neerincx[1,3], and Rosemarijn Looije[1]

[1] Netherlands Organization for Applied Scientific Research TNO, Kampweg 5, (3769 DE) Soesterberg
[2] Vrije Universiteit, Faculty of Sciences - dept. Computer Science, De Boelelaan 1081, (1081 HV) Amsterdam, The Netherlands
[3] Delft University of Technology, Faculty of Electrical Engineering, Mathematics, and Computer Science (EEMCS), P.O. Box 5031, (2600 GA) Delft, The Netherlands

Abstract. Based on a 'learning by playing' concept, a basic arithmetic learning task was extended with an engaging game to achieve long-term educational interaction for children. Personalization was added to this learning task, to further support the child's motivation and success in learning. In an experiment, twenty children (aged 9-10) interacted three times, spread over days, with a robot using the combined imitation and arithmetic game to test this support. Two versions of the robot were implemented. In one implementation the complexity of the arithmetic progressed towards a predefined group target. In the other version the assignments increased in complexity until a personal end level was reached. A subsequent free-choice period showed that children's motivation to play (and learn) was high, particularly when the game progressed to a personal target. Furthermore results show that most children in the last condition reach higher levels compared to the predefined group level.

1 Introduction

Children with a chronic lifestyle related disease have to take care of more aspects in daily life compared to their healthy peers. Support for these children in daily activities might therefore be beneficiary to them. The ALIZ-E project is aiming at a social robot for long-term interaction with these children. This robot should be able to perform three different roles over a relatively prolonged period of time: a *buddy* that provides a personalized and engaging interaction, an *educator* that teaches relevant knowledge and skills to 'empower' the child, and a *motivator* that persuades the child to adhere to a healthy lifestyle (e.g. the therapy, diet, medication) [10]. Several robot functions that support these roles are being developed incrementally, in an iterative process.

The overall scenario is based on a medical setting in Italy, where children diagnosed with diabetes will spend a complete week in the hospital after diagnosis to learn about the illness and the implications of it. For these (young) children one week away from home is a long time. The ALIZ-E robot intents to make

B. Mutlu et al. (Eds.): ICSR 2011, LNAI 7072, pp. 153–162, 2011.

the time in the hospital more pleasant, while supporting the education of the child's illness. Basic arithmetic skills help children with diabetes to count the carbohydrate intake for their nutrients. This ability can therefore contribute to young diabetics' self-efficacy.

In the project we use 'learning by playing' as a concept for the interaction. We combine the basic arithmetic learning task with an engaging game to achieve the project goal for long-term interaction. This last game is an imitation game, in which robot and child copy each others sequence of arm movements and, subsequently, add a new movement to this sequence. The robot attunes the number and repertoire of moves to child's performance based on principles of motivational feedback, in such a way that the children like to continue playing until they achieve their target [14]. This ensures that the child keeps being challenged which is an important factor in both intelligent tutoring systems [17] and game theory [6]. The imitation game balances the perceived challenges with the perceived skills of the child and proves to be challenging for the children. By additional personalization of the arithmetic assignments, we expect to further improve the child's motivation and learning performance. These effects are studied in an experiment.

How to measure motivation of young children is a non-trivial question. In addition to questionnaires and observations during the game, we will evaluate the motivation for interaction with the robot by providing a free-choice period [13] [16]. Furthermore, we choose to perform the experiment with healthy children, since we want to burden sick children as little as possible [5]. By ensuring that the general characteristics of the children in the experiment are similar to the diabetic target group (e.g. age), we expect to find principles that apply for both groups. In future experiments we plan to test this in a group of children with a chronic disease.

2 Aspects of Motivation

Literature distinguishes two types of motivation: intrinsic and extrinsic motivation. Our research objective is to establish long-term motivation, ultimately to make a change in behaviour possible. Extrinsic motivation, though effective for short-term task compliance, has been shown to be less effective than intrinsic motivation for long-term task compliance and behaviour change' [8]. We will therefore focus on *intrinsic* motivation. Fasola and Mataric [8] indicate several factors that contribute to intrinsic motivation, including praise, competition, real-time feedback of performance, optimal challenge, self-efficacy and self-determination. Vallerand et al. [16] describe several variables that decrease the intrinsic motivation and should therefore be avoided. These variables include: material rewards, surveillance, deadlines, lack of self-determination and negative performance feedback.

For this study, we used the imitation game. In this game, the robot and the child build sequences of arm movements together. Turn by turn the players repeat the existing sequence and add a new movement to the sequence. During the game, the robot gives motivational verbal feedback to maximize the performance of the child. Most motivational aspects were already incorporated into the imitation game: positive robot feedback (praise, real-time feedback on performance), no material reward for the child and the absence of deadlines or negative performance feedback. Other aspects were difficult to manipulate: competition, self-efficacy, self-determination. Optimal challenge is a aspect we have a closer look into: when a game is too easy, the player will become bored as opposed to the game being to difficult, which will result in the user becoming frustrated or anxious [8,6]. The optimal challenge is thus when there is a balance between perceived difficulty and perceived skills by the user. In the study presented here we implemented one version of arithmetic that should approximate the optimal challenge and one that does not.

3 Implementation

For this study we implemented and extended an imitation game with arithmetic assignments to mix fun and education. The new game is composed of two components: making arm movement sequences and solving arithmetic assignments. In line with Cohen et al. [2] emotional feedback in the game is attuned to match children's expectations. The game is presented to the children as a secret agent game, where arm movements are a secret code and to crack the code of the bad guys, the children have to solve arithmetic assignments. The performance on the two components of the game are not linked with each other.

Several worlds exist in the imitation game. The arm movements increase in difficulty depending on the current world (starting using one arm: 'left arm up' and extending towards both hands 'Both arms down'). Within a 'world', each player (the robot and the child) repeats the entire sequence and makes up a new movement, which is added to the sequence. The sequence ends when the length of the current world is reached or when the child makes a mistake. To prevent deception of the child, the robot does not make mistakes. The progress in 'worlds' is attuned to the performance of the child.

For the arithmetic implementation, 29 levels with 10 assignments are constructed. The levels have an increasing arithmetic difficulty (e.g. level 1: '$6 + 1$', level 10: '$41 - 10$', level 20: '4×42', level 29: '$1005 \div 67$'). The assignments are selected randomly within a level and displayed on a screen next to the robot (see Figure 1). The robot provides motivational verbal feedback after each answer.

Two versions of the robot game are implemented: one that has a predefined arithmetic group level as end goal and one in which children could reach the boundary of their arithmetic capabilities. Next to this distinction there is a difference in the learning algorithm between the two implementations. As long as no mistakes are made both versions have an increase of three levels each step, for fast convergence to an appropriate level. When a mistake is made in the group level version the level is increased by one from that moment on, resulting in a more moderate learning curve. For the personalized level implementation, a simple form of sensitivity analysis is used [15]. In the case of a mistake, the level is decreased by one, increasing the self-perception of arithmetic skills. Next to this, the levels are also increased by one from that moment on. Other motivational aspects including the arm movement part are not manipulated, in order to get a fair comparison between group and personalized level implementation.

The group goal is set at level 20, which was considered the appropriate level for children half way through year six (fourth grade in U.S.) based on information from Goffree [9] and Borghouts [1]. For the personalized level 29 is the maximum level, because the chances are slim that the children reach this level.

The robot contained a user model, which kept track of the movement progress and arithmetic level of each participant. The participant continues with the movements and assignments in the level they ended last interaction time.

4 Experimental Method

4.1 Participants

Participants were 20 Dutch children (11 F and 9 M, age 9 - 10 years) from elementary school 'Het Spoor' (Zeist, The Netherlands). This is a Jenaplan school, meaning that each child follows their own learning curve but still has to reach goals within a time frame. Balancing for their gender, the participants were randomly divided in two groups of 10 participants. All the parents/caregivers signed an informed consent.

4.2 Materials

NAO Robot. The robot used in this experiment was the NAO (Aldebaran Robotics, see Figure 2). The NAO was provided with a unisex name: Charlie. Charlie, and names with similar pronunciation, is an uncommon name in the Netherlands both for boys (494 in 2006) and girls (363 in 2006). We provided no clues about the gender of the robot, since we wanted to prevent the children being prejudiced to liking the robot because of its gender. Fluency TTS (v4.0, using neutral voice 'Diana') was used to generate the wav-files the NAO used. The software for executing the imitation game involves: a Wizard of Oz interface, a dialogue model, a user model, the arithmetic assignments database. The control software ran on various computers.

Wizard of Oz. In order to test the feasibility of components before complete implementation, we use a Wizard of Oz set-up. In this interface, the experimenter does the sensing (e.g. the wizard interprets the movements and speech of the children), initiates the dialog and controls the laptop that displays the assignments and the progress. The movements of the robot are also preprogrammed and the experimenter just has to press the right order of buttons to make the robot imitate the children. To the children it looks like the robot actually recognizes and remembers the set. At a later stage fully autonomous robot behaviours will be tested within the project.

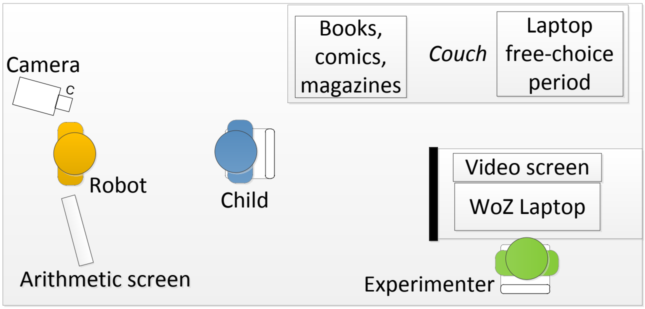


Fig. 1. The experimental setup

Fig. 2. The NAO robot

Experimental set-up. The experiment was conducted at the school in an office space. Unfortunately, there was no possibility for the experimenter to occupy a different room nearby, so the experimenter was in the same room as the child and the robot. The effects of the presence of the experimenter were minimized by placing a covering screen. The interaction was recorded on video. Figure 1 shows the experimental setup.

4.3 Experimental Design

The experiment performed had a between-subject design. The independent variable was the goal that could be reached in the arithmetic assignments, either the group or the personalized level. Between the two groups, one group interacted with a robot that adapted the level of the assignments to the child's performance and could proceed beyond the group level. The other group interacted with a group goal robot that followed a standard learning curve where the group level was the highest level that could be reached. The dependent variables were arithmetic performance and intrinsic motivation.

Measures. Two measures were used to measure the intrinsic motivation of the children to play the game with Charlie. A **questionnaire (subjective)** was constructed based on the Intrinsic Motivation Inventory (IMI)[1] (see other research [3][12][7]). Two of the seven subscales of the IMI were included: Interest/enjoyment (intrinsic motivation for playing game with the robot, 7 questions) and Relatedness (bond with the robot, 8 questions). The answers were measured using 7-point Likert scales (1 being negative and 7 positive). The original questionnaire in total and the separate subscales individually, were all validated. The questionnaire was translated into Dutch focused on children. The layout was altered for every session to keep children motivated to complete the questionnaire.

[1] http://www.psych.rochester.edu/SDT/measures/IMI_description.php

As an objective measure, the **free-choice period** [13] [16] was used. The free choice period was a period of five minutes in which the child could choose what to do: keep playing with the robot, read children's comics or do interactive Internet learning games on a computer. The time spend interacting with the robot was measured and functioned as an objective measure for the intrinsic motivation of the child to interact with the robot.

4.4 Procedure

During a short introduction in class, the children were able to see the robot beforehand. For the individual interaction moments, the experimenter introduced the child to the robot and explained the course of the experiment. Each interaction session lasted about 20 minutes, based on the average attention span for children of this age [4]. The child played the game with the robot for about 15 minutes. The game was ended after the 13th minute at a natural moment when the level was completed, resulting in a 13 to 17 minute interaction time. Afterwards the free-choice period was started by the experimenter. The researcher stated that the experiment had ended and that the child had 5 minutes to do as it pleased, choosing between the mentioned options as long as it stayed inside the room (options were presented in a random order). Finally, the child completed the questionnaire. The experiment was performed three times for each child over the course of two weeks. The rationale behind this was to experiment with the constructed user model and to overcome the initial enthusiastic response displayed by children when first meeting the robot. To reward the children, they received a picture of themselves with the robot. The school received technical Lego and was given a robotics lesson for the class after all sessions were completed.

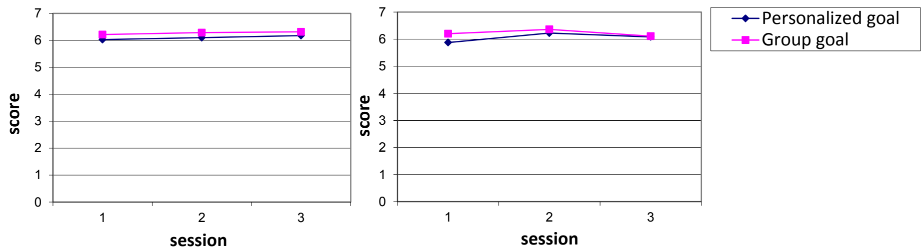
5 Results

5.1 Motivation

First the quantitative results will be discussed, based in the two different motivation measures used in the experiment.

Questionnaire. The results of the intrinsic motivation questionnaire are analyzed for each session. The answers represent the motivation to play the game with the robot and the bond with the robot. From the analysis the participant that did not start the third session is excluded and missing data is filled with a random participant from this condition to make ANOVAs possible.

We expected that the children in the personalized goal robot condition would score higher on the motivation scale than the children in the group goal condition. Results show that both scales are rated high (see Figure 3). The standard deviations are small, they ranged for Interest/Enjoyment from 0.39 to 0.70 and for Relatedness from 0.37 to 0.66. For Interest/Enjoyment the repeated measures ANOVA over the runs has as result: $F(2, 48) = 0.01$, $p=0.99$. The result



(a) Interest/Enjoyment **(b)** Relatedness

Fig. 3. Questionnaire results

Table 1. Results of the free-choice period. Time interacting with the robot (in mm:ss). * entails the child stopped the interaction before the free choice period started, ** entails the child was absent

Personalized goal robot				**Group goal robot**			
Child	**Run1**	**Run2**	**Run3**	**Child**	**Run1**	**Run2**	**Run3**
1	*	*	0:00	2	0:00	0:00	0:00
3	2:20	3:13	5:00	4	5:00	5:00	5:00
5	5:00	0:00	0:00	6	5:00	0:00	0:00
7	0:00	5:00	5:00	8	0:00	0:00	0:00
9	0:00	0:00	0:00	10	5:00	0:00	0:00
11	5:00	5:00	**	12	5:00	0:00	0:00
13	5:00	5:00	5:00	14	0:00	0:00	0:00
15	0:00	0:00	0:00	16	5:00	5:00	5:00
17	5:00	5:00	4:30	18	0:00	0:00	0:00
19	0:00	0:00	0:00	20	0:00	0:00	0:00
AVG	2.6	2.6	2.5	AVG	2.5	1.0	1.0

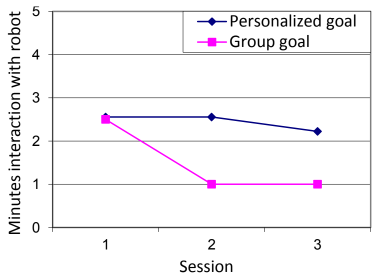
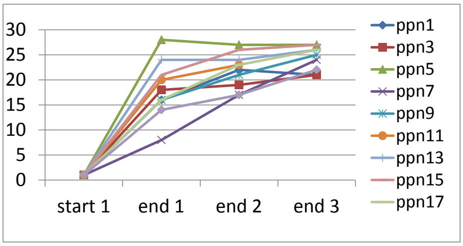


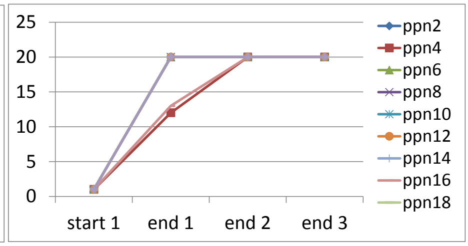
Fig. 4. Amount of time spend with robot during free-choice period

for the ANOVA for the Relatedness questionnaire is F(2, 48) = 0.16, p=0.85. Thus both questionnaires do not provide significant differences between the two conditions.

Free-choice period. Table 1 shows the amount of time spend with the robot per participant for each session during the free-choice period and Figure 4 shows the means graphically. In the free-choice period following the first interaction the time spend with the robot is about the same (mean personalized = 2.6min, mean goal = 2.5min). This was expected beforehand, due to the new experience of the interaction with the robot. Video footage shows that the children were in general very excited to play with the robot. After the first session, the results started to differentiate between the two conditions. Most children that interacted with the personalized goal robot continued to play with the robot during the free-choice period, whereas the children that interacted with the group goal robot displayed,



(a) Personalized goal



(b) Group goal

Fig. 5. Performance for arithmetic assignments

on average, a decline in the amount of time spend with the robot during the free-choice period. Child 1 stopped the experiment before the free-choice period and child 11 was absent during the third session resulting in missing data points.

Because the results are not evenly distributed, we ran a nonparametric Mann-Whitney U-test to establish whether the differences between the two conditions are significant. A one-tailed Mann-Whitney U-test showed that the difference was significant ($p < 0.05$) in favour of the personalized goal robot.

5.2 Results Arithmetic Aspects

We expected the children to reach arithmetic level 20, which corresponded with half way through 6th grade. However figures 5a and 5b show that the children that interacted with the personalized goal robot, performed above the expected norm on the arithmetic assignments (average 24.7). Especially child 5 stands out in arithmetic skills. The graphs show that most children already reached level 20 after the second interaction. From these results, we can derive that most children participating in the experiment are ahead in their arithmetic education and that playing with a personalized goal robot makes sense, since the individual levels differ from each other.

We looked into the interaction between the free-choice period and the performance on the arithmetic assignments. When Figure 5a and Figure 5b are linked with Table 1, it shows that the two children who played with the group goal robot during the free-choice period after session 2 and 3, were actually the children that did not reach level 20 after the first interaction. It appears as though the continuing increase in level motivated the children to play with the robot during the free-choice period. When looking at the children that played with the personalized goal robot, we see a similar trend. Child 5 performed very well on the assignments and reached his personal level during the first session. During the free-choice period the child chose to read instead of playing with the robot. However, Child 13 who also reached his personal level at the first session, did continue playing with the robot during the free-choice period. Hence, some children who perform at top level still like to play with the robot.

6 Conclusion and Discussion

In this paper we present a study that builds upon the principles of learning by playing. By combining a basic arithmetic task with an engaging game, we create a robot game for children. In an experiment we look whether personalization of the learning task has an effect on children's motivation and learning. In general we found that the children are very motivated to play the game with the robot. The motivation stays at a high level for all three interaction moments. The objective motivation, free-choice period, stays high when they interact with a robot offering a personalized learning goal. Most children who play with the personalized goal robot keep interacting with the robot the full five minutes of the free-choice period and the two children who are a bit slower to reach level 20 in the group goal session keep interacting with the robot during the free-choice period.

The personalized goal version shows that the group goal is not high enough for most of the children to reach their maximum capabilities. The group goal is thus not challenging. In sum, this robot game provides a promising approach to support long-term interaction even when the interaction is not all about fun. This is promising for the use of a social robot for long-term interaction with diabetic children. In a next study, diabetic children will participate in the study to see if the results can be reproduced with this specific population.

From a methodological perspective, the free-choice period proves to be very useful to study motivation effects with children. It appears that children answer the questions socially desirable. Despite several urges of the experimenter to rate how they really feel about the game, children seem to stay away from the 'negative' answers even though some children seem sometimes a little bored during the game (based on video footage). In future, we plan to include more detailed observations on communication behaviour, like eye-contact (gaze wondering off). In addition, we will improve the questionnaires. For example, research on Likert scales for children suggests to use a 3-point scale [11].

Acknowledgments. This work is (partially) funded by the European Union FP7 ALIZ-E project (grant number 248116). Furthermore the authors would like to thank the teachers and the children of 'Het Spoor' (the school) for their participation in this study.

References

1. Borghouts, C., Buter, A., Dekker, J., Hoogenberg, E., Kopmels, D., van Oostenbrugge, M.: Interactie in Rekenen. In: Bazalt Educatieve Uitgaven (2005)
2. Cohen, I., Looije, R., Neerincx, M.: Children's Recognition of Emotions in Robots' Face and Body. In: Proceedings of the 6th International Conference on Human Robot Interaction (HRI 2011), Lausanne, Switzerland, March 6-9 (2011)
3. Colombo, R., Pisano, F., Mazzone, A., Delconte, C., Micera, S., Carrozza, M.C., Dario, P., Minuco, G.: Design strategies to improve patient motivation during robot-aided rehabilitation. Journal of NeuroEngineering and Rehabilitation 4(3) (2007)

4. Cornish, D., Dukette, D.: The Essential 20: Twenty Components of an Excellent Health Care Team. Rosedog Pr (2009)
5. Coyle, D., Doherty, G., Matthews, M., Sharry, J.: Computers in talk-based mental health interventions. Interacting with Computers 19(4), 545–562 (2007)
6. Csikszentmihalyi, M., Abuhamdeh, S., Nakamura, J.: Handbook of Competence and Motivation. In: Flow, ch. 32, pp. 598–698. The Guilford Press, New York (2005)
7. Deci, E.L., Eghrari, H., Patrick, B.C., Leone, D.: Facilitating internalization: The Self-determination theory perspective. Journal of Personality 62, 119–142 (1994)
8. Fasola, J., Matarić, M.J.: Robot Motivator: Increasing User Enjoyment and Performance on a Physical/Cognitive Task. In: Proc. IEEE International Conference on Development and Learning (ICDL), Ann Arbor, MI, USA (2010)
9. Goffree, F., Oonk, W.: Reken Vaardig: op weg naar basale en professionele gecijferdheid. In: Noofdhoff Uitgevers B.V. (2004)
10. Looije, R., Neerincx, M., de Lange, V.: Children's responses and opinion on three bots that motivate, educate and play. Journal of Physical Agents 2, 13–20 (2008)
11. Markopoulos, P., Read, J., MacFarlane, S.: Evaluating children's interactive products: principles and practices for interaction designers. Morgan Kaufmann Pub. (2008)
12. Plant, R., Ryan, R.M.: Intrinsic motivation and the effect of self-consciousness, self-awareness, and ego-involvement: An investigation of internally-controlling styles. Journal of Personality 53, 435–449 (1985)
13. Reeve, J., Deci, E.L.: Elements of the Competitive Situation That Affect Intrinsic Motivation. Society for Personality and Social Psychology 22(1), 24–33 (1996)
14. Robben, S.M.B.: Graduation Thesis: It's NAO or Never! Facilitate Bonding Between a Child and a Social Robot: Exploring the Possibility of a Robot Adaptive to Personality (2011)
15. Russell, S., Norvig, P.: Artificial Intelligence, a modern approach. Pearson Education, New Jersey (2003)
16. Vallerand, R., Gauvin, L., Halliwell, W.: Negative Effects of Competition on Children's Intrinsic Motivation. The Journal of Social Psychology 126(5), 649–657 (1986)
17. Woolf, B.: Building intelligent interactive tutors, vol. 3. Elsevier (2007)

Attitude towards Robots Depends on Interaction But Not on Anticipatory Behaviour

Raymond H. Cuijpers*, Maarten T. Bruna, Jaap R.C. Ham, and Elena Torta

Eindhoven University of Technology,
Den Dolech 2, P.O. Box 513, 5600MB Eindhoven, The Netherlands
{r.h.cuijpers,j.r.c.ham,e.torta}@tue.nl, maarten_bruna@hotmail.com
http://hti.ieis.tue.nl/

Abstract. The care robot of the future should be able to navigate in domestic environments and perform meaningful tasks. Presumably, a robot that moves and interacts more intelligently gains more trust, is liked more and appears more humanlike. Here we test in three scenarios of differing urgency whether anticipatory walking behaviour of a robot is appreciated as more intelligent and whether this results in a more positive attitude towards the robot. We find no effect of walking behaviour and a main effect of urgency of the scenarios on perceived intelligence and on appropriateness. We interpret these results as that the type of interaction determines perceived intelligence and the attitude towards robots, but the degree of anticipation has no significant effect.

Keywords: cognitive robotics, social robotics, navigation, learning, anticipation.

1 Introduction

One important aspect in the field of social robots is navigating towards a person [7,18]. However, approaching a person introduces new problems: people are not stationary, a domestic environment is dynamically changing and cluttered, and people have plans, goals and intentions of their own[1,11]. Typically, this involves extracting the user's movement patterns [14,15]. Several methods for movement prediction have been developed [9,4], most of which use clustering of movement patterns to predict movement goals. For example, Bennewitz et al. [4] present an implementation of clustering using the expectation-maximisation algorithm. An important property of clustering-based approaches is that they use only low-level information (e.g. trajectory data) to predict the goal of the movement, which makes them sensitive to dynamically changing environments. Clearly, if a robot knew a person's intentions it could anticipate on the person's destination and move to an appropriate location in the appropriate time window[6,18]. A problem that is rarely addressed is that different people have different personalities. Thus, even if a robot is optimally capable of inferring one person's behaviour, it might

* Corresponding author.

B. Mutlu et al. (Eds.): ICSR 2011, LNAI 7072, pp. 163–172, 2011.

be suboptimal for another person. A solution would be that a robot can learn dynamically in real time so it is able to adapt to different persons through on-line learning [16].

Another way to look at this problem of robot navigation towards persons is to ask what the benefit would be if a robot were able to correctly infer a user's intentions. Does learning of high-level information about a person's goals improve the robustness and flexibility of a robot's navigational skills in such a way that the robot's behaviour appears more natural, trustworthy and socially intelligent to users? If it does, this would clearly be beneficial in care applications where robots assist people [12,8,10,5]. There are several factors that influence the users' opinion of a robot and the interaction with it. For example, Bartneck et al. [3] show that the appearance of the robot has a significant effect on its perceived intelligence. Dautenhahn et al.[7] have researched the preferred way for a robot to approach a seated human and found that people preferred an approach from the left or right side over a frontal approach. Althaus et al. [1] let a robot approach a group of humans in a similar way as humans do. They find that people rate the robot's behaviour as very natural, which increases its perceived intelligence. Pacchierotti et al. [17] show that a robot that shows anticipatory behaviour to avoid collision is rated positively. Green and colleagues [13,12] show that human-robot interaction also benefits from additional information for the user through augmented reality. Unfortunately, augmented reality involves additional equipment and training, which is undesirable in care applications and ambient assisted living. These studies confirm that natural, human-like behaviours help improve human-robot interaction from the user's point of view. The problem is to know which behaviour will appear as natural because it is highly dependent on the specific context in which it occurs. It is easy to imagine situations in which a robot that follows you around all the time is annoying, whereas it is perceived as useful in others.

In this study we look at whether anticipatory walking behaviour of a robot is perceived as more intelligent, and as a consequence, appears more trustworthy and likeable. For that purpose, we compare following behaviour, which is purely reactive without anticipation, interception behaviour, which extrapolates in time the person's walking trajectory, and walking to the anticipated destination, which requires inferring the goal of a person. It is conceivable that the degree with which these different behaviours are beneficial depends on the particular context. We focused on three concrete scenarios from the KSERA project (http://www.ksera-project.eu), which are based on the user needs of COPD patients and older persons in general. The three different scenarios are: phone call, medical emergency, and health exercise. These scenarios are chosen because of they are conceptually very different, in particular in their degree of urgency. We expected that a health exercise, which is not urgent, is most compatible with waiting at the anticipated location, whereas a medical emergency, which is very urgent, is most compatible with following behaviour, as the robot is always nearby. Thus, we expect the scenarios and behaviours to interact.

2 Methods

2.1 Experimental Setup

The robot used for the experiment was Nao (Aldebaran Robotics, France). Nao is a 58 cm tall fully programmable, humanoid robot with 25 degrees of freedom and is equipped with multiple sensors, multiple communication ports and an inbuilt Linux computer (Fig. 1a). The robot was manually controlled by the experimenter from another room through its inbuilt WiFi connection and inbuilt walking behaviours. The experiment took place in a living room setting; the layout of the room is shown in Fig. 1b). The set-up consisted of a dinner table (labelled by A), a coffee table (labelled by B) and a comfortable chair (labelled by C).

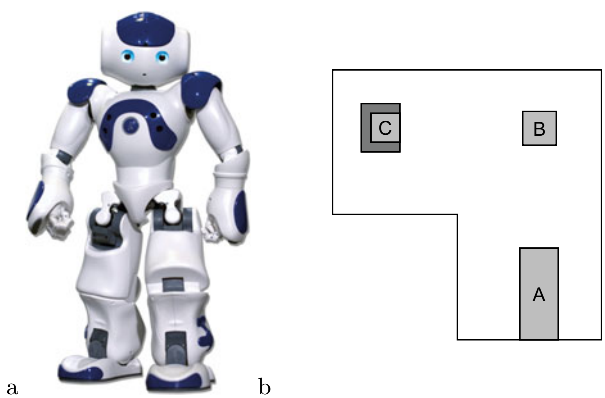


Fig. 1. a) The Nao robot of Aldebaran Robotics b) Layout of the experimental room equipped with a dinner table (A), a coffee table (B) and a comfortable chair (C)

2.2 Participants and Task

Fourteen participants (11 male, 3 female) whose age varied between 19 and 32 (mean age 24.5, SD = 4.3) participated in the experiment. Participants were unfamiliar with the research topic. They were all students or employees of the Eindhoven University of Technology. During the experiment, participants had to perform a task which required their interaction with a robot after which they had to fill out a questionnaire.

2.3 Design

The experiment followed a 3x3 design, with 3 ways of approaching by the robot and 3 different scenarios. The order of presentation was randomised and counter-balanced across participants. The three approaching behaviours were as follows:

1) Following - Nao follows the person. He announces his behaviour by telling the participant “I will follow you”. 2) Intercepting - Nao intercepts the person halfway between the coffee table and the chair. 3) Anticipation - Nao goes to the anticipated destination (the chair) and waits for the participant to arrive there. He announces his behaviour by telling the participant: “I think you will go to your chair". The three ways of approaching result in three different robot trajectories which are shown in Fig. 2.

To see whether the user evaluation of approaching behaviour depends on context three scenarios were devised based on use cases for the KSERA project (http://www.ksera-project.eu). The scenarios tested during the experiment were: 1) Phone Call - the robot approaches the participant, and says “There is a phone call for you” and starts making a ringing sound. The participant is instructed to reject the phone call by touching the robot’s head. 2) Medical Emergency - the participant is told to imagine that he or she feels ill and has to wait at the coffee table for the robot to arrive there. After arriving close by and in front of the participant the robot says “Do not worry I will call the doctor”. 3) Health exercise - Nao approaches the participant and tells him/her “this might be a good time to exercise”. Since we were only interested in the approaching behaviour, the robot stopped acting after delivering its message or after its touch button on the head was touched. This also ensures that, apart from the conceptual difference and the user’s mind set, the robot behaviours are very similar across scenarios.

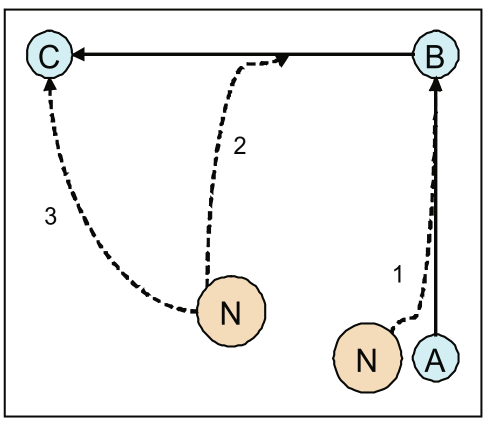


Fig. 2. The different approaching behaviours of the robot from two starting locations (N) as the user goes from $A \rightarrow B \rightarrow C$: 1) Following the user 2) Intercepting the user 3) Waiting for the user at the anticipated location. The starting locations are chosen in such a way that the walking duration is similar across conditions.

2.4 Procedure

After arriving in the lab, each participant was asked to sit down at the coffee table and read and sign an informed consent form. When they had read it and had no further questions the experimenter moved next door and controlled the robot from there. The rooms were separated with a one-way mirror so that the experimenter could watch the participants but could not be seen himself. At the beginning of each trial participants received instructions on paper at the dinner table (Fig. 1b). They were told that they were to imagine that they had just finished a meal at the dinner table and wanted to read the paper in their favourite chair. To do this they first had to pick up the paper at the coffee table (B in Fig. 1b). On the coffee table there was a laptop which showed new instructions and listened to a 30-second audio clip from the BBC Radio news. This was necessary because Nao walks very slowly, so that participants had to be slowed down to enable Nao to keep up with them. Afterwards participants received a question about the radio clip in order to make sure participants paid attention and to give the impression that it was part of the task. The answers themselves were not used for analysis. After the audio clip had finished they were instructed either to pick up the paper and walk to the chair (C in Fig. 1b) or stay at the coffee table because they supposedly felt ill and had to wait for the Nao robot to come to their aid. When the participants had reached the chair or Nao had reached the participant the trial was over and participants had to rate the robot's behaviour during that interaction by filling out a questionnaire (see section 2.5). The questionnaire also contained a question about the news clip in order to make sure the participants paid attention to the news clip. After filling out the questionnaire, the next trial started. Each participant performed 9 trials in total. At the end of the experiment participants were asked for comments or remarks, the research topic was explained, they received their financial reward and were thanked for their participation.

2.5 Questionnaire

The questionnaire was based on the Godspeed questionnaire by Bartneck and colleagues [2]. This is questionnaire is especially developed for assessing whether people think of the robot as a liveless, unintelligent machine or a social, intelligent agent. The questionnaire measures 5 dimensions of human-robot interaction: anthropomorphism, animacy, likeability, perceived intelligence and perceived safety. As an additional dimension we added whether people felt the robot's behaviour was appropriate given the circumstances. Apart from the latter all dimensions were evaluated on 5 or 6 items. Each item was evaluated using a 5 point Likert scale. For example, the item Fake-Natural could be indicated by very fake, fake, neutral, natural, and very natural. The dimensions and items are listed below:

- Anthropomorphism: The extent to which the robot is "humanlike".
  - Fake - Natural
  - Machinelike - Humanlike
  - Unconscious - Conscious
  - Artificial - Lifelike
  - Moving rigidly - Moving elegantly
- Perceived Intelligence:
  - Incompetent - Competent,
  - Ignorant - Knowledgeable
  - Irresponsible - Responsible
  - Unintelligent - Intelligent
  - Foolish - Sensible
- Likeability:
  - Dislike - Like
  - Unfriendly - Friendly
  - Unkind - Kind
  - Unpleasant - Pleasant
  - Awful - Nice
- Animacy: The extent to which the robot is "alive".
  - Dead - Alive
  - Stagnant - Lively
  - Mechanical - Organic
  - Artificial - Lifelike
  - Inert - Interactive
  - Apathetic - Responsive
- Perceived Safety: How safe participants felt during the interaction.
  - Anxious - Relaxed
  - Agitated - Calm
  - Quiescent - Surprised
  - Unsafe - Safe
- Appropriateness: How appropriate is the robot's behaviour.
  - Inappropriate - Appropriate

2.6 Data Analysis

To test the effects of our manipulations on the different dimensions of the questionnaire we will use a repeated measures analysis of variance (ANOVA) with the Likert scale scores as dependent variable and both manipulations (approaching behaviour and scenario) as independent variables. Before using the data to compare the effects of the experimental manipulations, the consistency and reliability of the items of each scale was tested. Items that had a strong detrimental effect on consistency reliability were removed. To test this we computed the Cronbach's alpha statistic. Items are considered to have good consistency if Cronbach's alpha exceeds 0.7. We only found one moderate reliability which occurred for the perceived safety dimension ($\alpha = 0.647$). This could be improved by removing the quiescent – surprised item. In the Table 2.6 below the resulting number of items and the computed Cronbach's α are given for each dimension.

Table 1. Internal consistency reliability scores

	Items	Cronbach's α
Anthropomorphism	5	0,817
Animacy	6	0,806
Likeability	5	0,883
Perceived Intelligence	5	0,870
Perceived Safety	3	0,831

3 Results

The mean Likert scores are quite high. In fact, the average across all behaviours and scenarios is ≥ 3.50 for all dimensions. This shows that participants in general had a positive attitude towards the Nao robot and the behaviours and scenarios that were used.

In Fig. 3 the mean scores of the Likert scales are shown for each dimension. Each panel shows the mean scores for a single dimension and for each scenario (x-labels) and robot behaviour (symbols). When comparing the different walking behaviours within a given scenario and dimension, it is clear that in most cases there seems to be no effect of walking behaviour. Indeed, a repeated measures ANOVA reveals that there is no significant main effect walking behaviour for each of the dimensions (Animacy: $F(2, 24) = 2.15, p = 0.14$; Anthropomorphism: $F(2, 24) =< 1, p = 0.45$; Likeability: $F(2, 24) = 1.54, p = 0.24$; Perceived Intelligence: $F(2, 24) = 1.81, p = 0.18$, Perceived Safety: $F(2, 24) = 1.22, p = 0.31$, Appropriateness: $F(2, 24) = 1.22, p = 0.76$). On the other hand, when comparing between scenarios the 'Exercise' scenario tends to score lower than 'Medical emergency' for the dimensions (Perceived Intelligence, Perceived Safety, and Appropriateness). The repeated measures ANOVA confirms that there is a main effect of scenario for Likeability ($F(2, 24) = 3.43, p = 0.049$), Perceived Intelligence ($F(2, 24) = 7.01, p = 0.04$), Perceived Safety ($F(2, 24) = 4.34, p = 0.025$) and Appropriateness ($F(2, 24) = 6.76, p = 0.05$), but not for Animacy ($F(2, 24) = 2.298, p = 0.12$) and Anthropomorphism ($F(2, 24) = 1.29, p = 0.30$). Pairwise comparison with Bonferroni correction revealed significant difference between medical emergency and exercise for Perceived Intelligence ($p = 0.04$) and Appropriateness ($p = 0.03$). We also checked whether there were interaction effects between scenarios and robot behaviours, but none were significant.

4 Discussion

To investigate how a person's attitude changes towards a robot depending on the anticipatory character of its walking behaviour and the urgency of particular context, we performed a study where we had subjects rate the robot's behaviour based on the Godspeed questionnaire developed by Bartneck et al. [2]. We expected that people would have a more positive attitude to robots the anticipatory their walking behaviour is, but that this effect would be moderated or even reversed depending on the context within which this behaviour takes place. We used three scenarios based on the user needs of elderly and COPD patients obtained in the KSERA project. The scenarios differed in urgency with which the robot should act. Apart from this the only difference between scenarios is the message delivered by the robot and whether or not the participant had to touch the robot's head. Results indicated that the walking behaviour did not affect our participants' ratings. We only found a main effect of the scenarios. Medical emergency was rated as more positive than health exercise in terms of appropriateness and in terms of the intelligence of the robot as perceived by users.

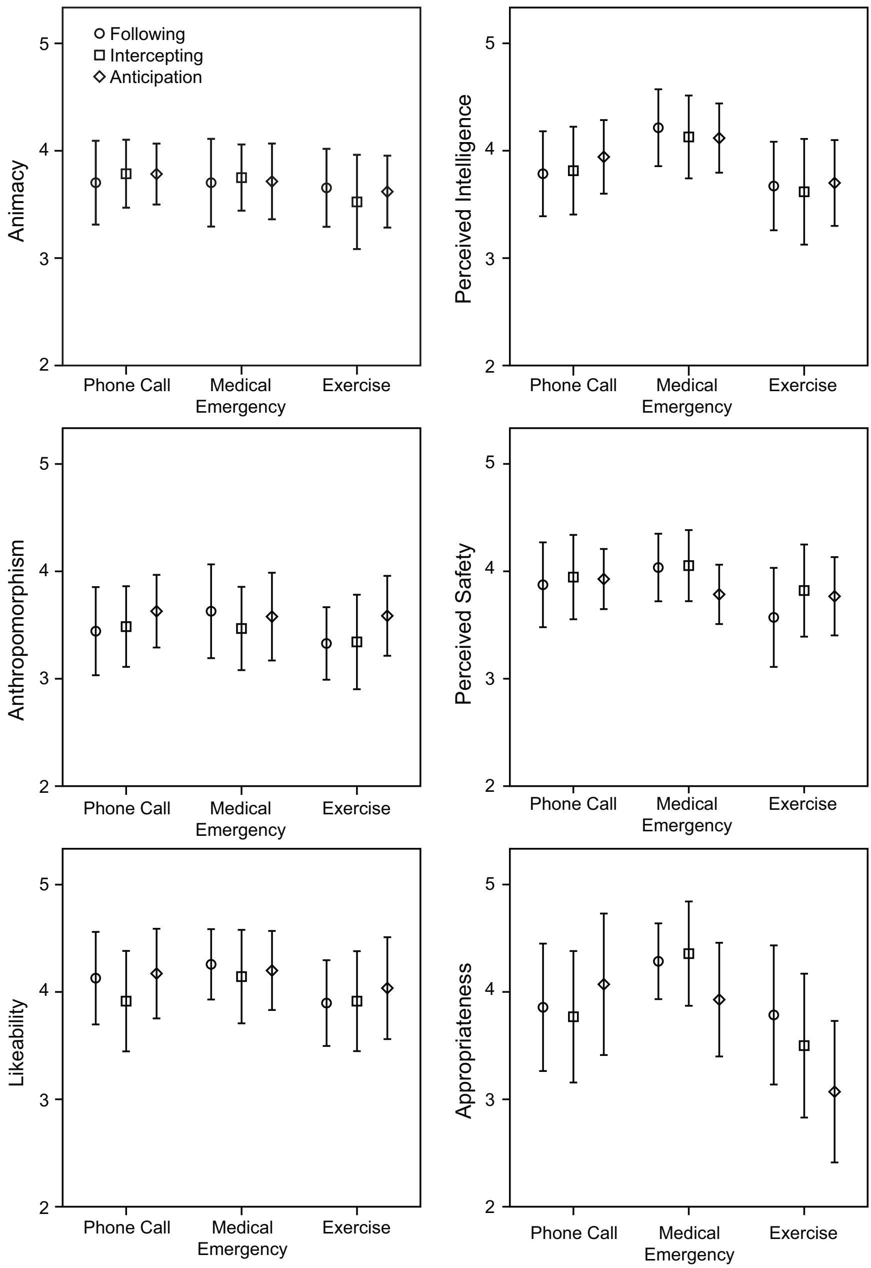


Fig. 3. Each panel shows the average Likert scale score for a single dimension of the Godspeed questionnaire. The labels on the x-axis indicate the scenarios and the symbols indicate the robot's walking behaviours. The error bars indicate 95% confidence intervals.

A possible confound could be the difference in interaction styles between the scenarios. In the health exercise condition the robot only talks to the person, in the phone call condition the user also 'answers' the phone by touching the robot, and in the medical emergency condition the robot comes to aid the user. In other words, it may not be the urgency but the degree of interaction that plays an important role. Contrary to our expectation we found no interaction effects. Intuitively, a robot that waits at the anticipated destination is rated more positively only when this does not interfere with being nearby. Consequently, we expected that waiting at the anticipated location would be rated more positive in the health exercise scenario, which is least urgent, than in the medical emergency scenario, which is most urgent. This is not what we found.

The current study had a number of practical limitations, which could explain in hindsight why there was no effect of walking behaviour. The Nao cannot walk fast enough to intercept anyone even when he or she has a slow pace. Therefore we introduced a secondary task to delay the participant. This worked well as the robot arrived at the right location at the right time, but it may also have distracted the participants too much. Another explanation is that there was only one possible destination for the participants. Therefore it is possible that the participant did not perceive the anticipatory behaviour as anticipatory, even though the robot verbally expressed its behaviour. The duration and sophistication of human-robot interaction was limited to keep the experiment duration manageable and to keep learning effects as limited as possible. In addition it makes the scenarios comparable. However, it might be the case that the duration of the human-robot interaction was too short to significantly change a person's attitude towards the robot. In future experiments these problems could be remedied by using multiple destinations and more elaborate interactions with the robot.

In conclusion, we found that the attitude towards robots does not depend on its anticipatory behaviour, but on the type of interaction and the urgency of the context within which this interaction takes place.

Acknowledgements. The research leading to these results is part of the KSERA project (http://www.ksera-project.eu) and has received funding from the European Commission under the 7th Framework Programme (FP7) for Research and Technological Development under grant agreement n° 2010-248085.

References

1. Althaus, P., Ishiguro, H., Kanda, T., Miyashita, T., Christensen, H.I.: Navigation for Human-Robot Interaction Tasks. In: Proceedings of the IEEE International Conference on Robotics and Automation, pp. 1894–1900 (2004)
2. Bartneck, C., Kulić, D., Croft, E., Zoghbi, S.: Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. International Journal of Social Robotics 1(1), 71–81 (2009), doi:10.1007/s12369-008-0001-3
3. Bartneck, C., Kanda, T., Mubin, O., Al Mahmud, A.: Does the Design of a Robot Influence Its Animacy and Perceived Intelligence? International Journal of Social Robotics 1(2), 195–204 (2009)

4. Bennewitz, M., Cielniak, C., Burgard, W., Thrun, S.: Learning Motion Patterns of People for Compliant Robot Motion. The International Journal of Robotics Research 24(1), 31–48 (2005)
5. Breazeal, C.: Toward sociable robots. Robotics and Autonomous Systems 42(3-4), 167–175 (2003), doi:16/S0921-8890(02)00373-1
6. Cuijpers, R.H., Schie, H.T., van Koppen, M., Erlhagen, W., Bekkering, H.: Goals and means in action observation: A computational approach. Neural Networks 19(3), 311–322 (2006)
7. Dautenhahn, K., Walters, M.L., Woods, S., Koay, K.L., Nehaniv, C.L., Sisbot, A., Alami, R., Siméon, T.: How May I Serve You?: A Robot Companion Approaching a Seated Person in a Helping Context. In: Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-robot Interaction, pp. 172–179 (2006)
8. Dautenhahn, K.: Socially intelligent robots: dimensions of human–robot interaction. Philosophical Transactions of the Royal Society B: Biological Sciences 362(1480), 679–704 (2007), doi:10.1098/rstb.2006.2004
9. Foka, A.F., Trahanias, P.E.: Probabilistic Autonomous Robot Navigation in Dynamic Environments with Human Motion Prediction. International Journal of Social Robotics 2, 79–94 (2010)
10. Fong, T., Nourbakhsh, I., Dautenhahn, K.: A survey of socially interactive robots. Robotics and Autonomous Systems 42(3-4), 143–166 (2003), doi:16/S0921-8890(02)00372-X
11. Yamaoka, F., Kanda, T., Ishiguro, H., Hagita, N.: A Model of Proximity Control for Information-Presenting Robots. In: Proceedings of the 5st ACM SIGCHI/SIGART Conference on Human-robot Interaction - HRI 2010, vol. 26(1), pp. 187–195 (2010)
12. Green, S.A., Billinghurst, M., Chen, X., Chase, G.J.: Human-Robot Collaboration: A Literature Review and Augmented Reality Approach in Design. International Journal of Advanced Robotic Systems 5(1), 1–18 (2008)
13. Green, S.A., Chase, J.G., Chen, X., Billinghurst, M.: An Augmented Reality Human-Robot Collaboration System. In: Chen, X.Q., Chen, Y.Q., Chase, J.G. (eds.) Mobile Robots - State of the Art in Land, Sea, Air, and Collaborative Missions (2009) ISBN: 978-953-307-001-8 (in Tech.)
14. Kanda, T., Glas, D.F., Shiomi, M., Hagita, N.: Abstracting People's Trajectories for Social Robots to Proactively Approach Customers. IEEE Transactions on Robotics 24, 1382–1396 (2009)
15. Kasteren, T., van Noulas, A.K., Englebienne, G., Kröse, B.J.A.: Accurate Activity Recognition in a Home Setting. In: Proceedings of the 10th International Conference on Ubiquitous Computing, pp. 1–9 (2008)
16. Nguyen, N.T., Phung, D.Q., Venkatesh, S., Bui, H.: Learning and Detecting Activities from Movement Trajectories Using the Hierarchical Hidden Markov Model. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), vol. 2, pp. 955–960 (2005)
17. Pacchierotti, E., Christensen, H.I., Jensfelt, P.: Human-robot embodied interaction in hallway settings: a pilot user study. In: IEEE International Workshop on Robot and Human Interactive Communication, ROMAN 2005, pp. 164–171 (2005), doi:10.1109/ROMAN.2005.1513774
18. Torta, E., Cuijpers, R.H., Juola, J.F.: A model of the user's proximity for Bayesian inference. In: Proceedings of the 6th International Conference on Human-robot Interaction - HRI 2011, p. 273 (2011), doi:10.1145/1957656.1957767

Listening to Sad Music While Seeing a Happy Robot Face

Jiaming Zhang and Amanda J.C. Sharkey

Neurocomputing and Robotics Group,
Department of Computer Science,
University of Sheffield,
Regent Court, Portobello, Sheffield S1 4DP, UK
acq08jz@sheffield.ac.uk, amanda@dcs.shef.ac.uk

Abstract. Researchers have shown it is possible to develop robots that can produce recognizable emotional facial expressions [1, 2]. However, although human emotional expressions are known to be influenced by the surrounding context [7], there has been little research into the effect of context on the recognition of robot emotional expressions. The experiment reported here demonstrates that classical music can affect judgments of a robot's emotional facial expressions. Different judgments were made depending on whether the music was emotionally congruent or incongruent with the robot's expressions. A robot head produced sequences of expressions that were designed to demonstrate positive or negative emotions. The expressions were more likely to be recognized as intended when they occurred with music of a similar valence. Interestingly, it was observed that the robot face also influenced judgments about the classical music. Design implications for believable emotional robots are drawn.

Keywords: robot, emotions, facial expressions, surrounding context.

1 Introduction

Suppose you were on your way to a museum that had a robot receptionist. On your journey, you listen to a very sad piece of classical music. When you enter the museum, the robot receptionist greets you with a tentative smile, just as a news broadcast comes over the speakers announcing that the football club you hate has just won the championship. When you are asked what the robot receptionist is feeling you answer "sad". Why? The music and football news have influenced your view of the robot's expressions. In the future, it is quite likely that we will encounter some real-life situations where a robot that produces emotional expressions will be surrounded by unexpected contradictory emotional contexts.

Previous work on creating robots that can make convincing emotional expressions has concentrated on the quality of those expressions, and on assessing people's ability to recognize them. For instance, Goris et al. report results that show how well people are able to recognize the emotional expressions of the Probo robot [1]. Breazeal [2] reports similar findings for Kismet. Such evaluations are useful, and show that it is

B. Mutlu et al. (Eds.): ICSR 2011, LNAI 7072, pp. 173–182, 2011.

possible to develop robots that can form recognizable emotional expressions. Both studies used the FACS (Facial Action Coding System) [3, 4] and Russell's circumplex model of affect [5, 6].

The FACS was initially developed as a set of guidelines for recognizing the facial expressions of humans but has been found to be a reliable tool for creating believable versions of six distinct and universal facial expressions (happiness, fear, surprise, anger, disgust/contempt, and sadness) for some emotional robots, even though some emotions (fear, and happy+surprised) were harder to make convincing [1, 2]. Russell's circumplex model of affect was used to create an emotion space for Kismet [5], and a new version used for Probo [6], and good recognition rates for their emotional expressions were obtained [1,2].

However, emotional expressions do not occur in a vacuum. Research on the recognition of human emotions [7] has shown that people's interpretation of human emotional expressions can be affected by the surrounding emotional context. We interpret a smile differently if we know that the person smiling has just heard that they have won the lottery, than we would if we knew that the same person had just been accused of bullying. Research also shows that such contextual effects are stronger if the emotional expression itself is a little ambiguous [7].

The facial dominance account of human emotions supported by Izard and Ekman [8, 9] assumed that a clear prototypical facial expression of a basic emotion will override any expectations derived from the surrounding situation. However, an increasing amount of evidence supports an account of limited situational dominance, whereby facial expressions are seen as only one of the elements contributing to the emotion attributed to a person, and the surrounding situation is found to have a strong effect. For example, Carroll and Russell [10] report a study in which they found that even when facial expressions were prototypical of an emotion, the context in which they occurred determined the specific emotion believed to be expressed by the expresser. In their study, the surrounding situation affected the interpretation of the facial expression, but sometimes the effects are bidirectional. For instance, when De Gelder and Vroomen [11] explored the combination of information from human facial expressions and voice tone, they found bidirectional contextual effects. They presented a series of pictures of faces ranging from happy to sad, accompanied by a sad voice, a happy voice or no voice. When the emotions matched, people's reaction time in judgments was faster than when they were mismatched. The vocally expressed emotions affected judgments of the facial emotions, and vice versa.

If context affects our interpretation of human faces, it seems likely that it will also affect our interpretation of robot expressions. Most research on robot emotional expressions has not looked at the effects of context, but the question has been explored in some avatar research [12, 13, 14, 15, 16]. Mower et al. [15, 16] used computer simulated avatars in an effort to determine how participants made emotional decisions when presented with both conflicting (e.g. angry avatar face, happy avatar voice) and congruent information (e.g. happy avatar face, happy avatar voice) in an animated display consisting of two channels: the facial expression and the vocal expression. They found that: (1) when they were presented with a congruent combination of audio and visual information, users could differentiate between happy and angry emotional expressions to a greater degree than when they were presented with either of the two channels individually; (2) when faced with a conflicting emotional presentation, users predominantly attended to the vocal channel rather than the visual channel. In other words, they were more likely to see the avatar's expressed emotions as being expressed by means of its voice, than reflected in its facial expressions. Their findings indicate that the observer's recognition of facial expressions can be influenced by the surrounding context, and also that emotion conveyed by other modalities, in this case the voice, can override that expressed by the face.

Similar results were found by Hong et al. [12] who adapted neural networks to map audio tracks (three sentences, namely "It is normal.", "It is good." And "It is bad." were read in a neutral tone without emotions) to facial expressions (neutral, smile, or sad) of a real-time speech-driven synthetic talking head. The audio tracks were either congruent or incongruent with the facial stimuli of an avatar, or a real human face. The study suggested that congruent audio stimuli improved subjects' ability to recognize the facial expressions of the avatar and the real face. Noël et al. [13], on the other hand, found that recognition of the emotional expressions of an avatar was not improved by the presentation of congruent, rather than incongruent emotional text. However the texts they used were very short, and their emotional expressions were preselected on the basis of being extremely recognizable.

In a recent study by Zhang and Sharkey [17], evidence was found of the effect of context on the recognition of robot expressions. A moving robot head displayed a sequence of emotional expressions that could be described as either positive, or negative, and did so at the same time as an accompanying recording of either positive or negative BBC news. It was found that people were much better at recognizing the robot's expressions as intended when the valence of the expressions matched the valence of the news, than when the two were in conflict (e.g. positive robot expressions, and negative news and vice versa). Similar effects were also found when the robot was viewed in an emotional context that consisted of affective pictures selected from [18], which again either matched, or conflicted with the robot's expressions. It was also found that the recorded BBC news had a more dominant effect on emotional judgments than the robot's expressions, and that the affective pictures showed a weaker, but still dominant effect. This study showed that context can affect the recognition of robot emotional expressions, but represents a first stab at the issue. Questions remain about: (1) what kinds of context are likely to affect the recognition of robot emotional expressions, and about the circumstances under which such contextual effects are likely to occur; (2) if such effects occurred, are they always dominant or could they be bidirectional such that the robot emotional expressions can reversely affect the recognition of the surrounding contexts?

A robot's facial expressions can be viewed as a modality containing emotional information. At the same time, a surrounding context, such as a piece of classical music that is either happy or sad, can be treated as another modality. The two modalities may reflect congruent or incongruent emotions. In the present paper, we report an experiment designed to discover whether happy or sad classical music would also influence people's interpretation of a robot's emotional expressions. This experiment extends our earlier work [17] which only examined two types of context (pictures and news recordings). Previous research on the effects of context on the recognition of human, or avatar, faces has usually paired voices and faces together [10, 11, 12, 14, 15, 16]. It would be interesting to see whether other forms of emotional context would have a similar effect. In many ways, it makes sense that people integrate the emotional cues provided by voices, and faces. It would be rather more surprising if the emotional valence of accompanying music was found to have an effect.

We report an experiment in which a robotic head displayed a sequence of synthetic emotional expressions based on the principles of the FACS system [4]. These expressions either matched, or conflicted with, the emotional valence of accompanying classical music. The effect of this musical context on the recognition of the robot's expressions was investigated, as well as the extent to which the robot's expressions, or the music, had a stronger influence.

2 Hypotheses

A surrounding context can be either congruent or incongruent with the emotional valence of the simulated emotions shown by a robot. Emotional congruence has been shown to affect judgments on the emotions displayed by computer simulated avatars [15, 16] and on human faces [7]. In order to explore whether this is also the case when they make judgements about robot emotional facial expressions that are accompanied by happy or sad music, we formulated the primary hypothesis as follows:

Hypothesis 1 (H1): When there is a surrounding musical context, people will be better at recognizing robot emotions when that context is congruent with the emotional valence of the robot emotions than when the context is incongruent with the emotional valence of the robot emotions.

It was previously found in our study [17] that users were influenced more by the surrounding contexts than by robot's facial expressions when making judgments about the robot's emotional state. It is possible then that a musical context will not only influence the interpretation of the synthetic robot facial expressions, but even dominate that interpretation. Therefore, the validation of H1 leads to the following hypothesis:

Hypothesis 2 (H2): When subjects are presented with conflicting information from the robot's face and an accompanying emotional context, their perception of the robot emotions will be more influenced by the surrounding context than the robot itself.

3 Method

3.1 Interaction Design

We conducted a between-subjects experiment. This experiment was based on a robot head known as CIM. As in our earlier study [17], the FACS (Facial Action Coding System) [4] was applied to set up the parameters of the servos (for more details, please refer to [17]) to make the robot head produce sequences of the six static facial expressions (joy (e.g., AUs 6+12), fear (e.g., AUs 1+2+4+10+12), surprise (e.g., AUs

1+2+10+12+58+63), anger (e.g., AUs 2+23+42+44), disgust (e.g., AUs 4+9+15+16), and sadness (e.g., AUs 1+4+6+15+64)) shown below:

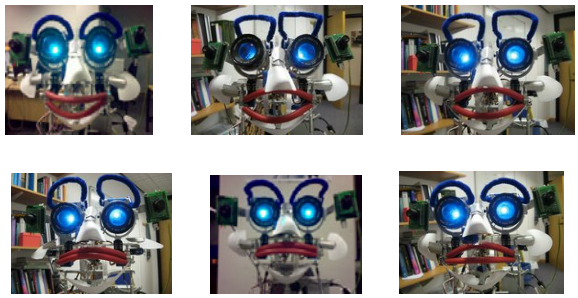


Fig. 1. Joy (top left), Surprise (top middle), Fear (top right), Sadness (bottom left), Anger (bottom middle), and Disgust (bottom right) of CIM. Note that the robot head neither have lid nor have jaw resulting in no AU 7 or AU 26, for instance, in use.

Two sequences of simulated facial expressions were developed to be shown together with the emotional contexts (the same sequences were used in [17]). They consisted of two types: Positive Affect, which mainly consisted of three different versions of joyful and surprised expressions (motions such as looking around and nodding were added in the gaps between these joyful and surprised expressions), and Negative Affect which mainly consisted of three different versions of sad, angry, and disgusted expressions (motions such as shaking and denying were added in the gaps between these sad, angry, and disgusted expressions). Each sequence of facial expressions was about three minutes long: the same time length as the classical music.

The 3 minutes long classical music used in the experiment was chosen from a list in provided in [19]. Subjects listened to two types of the classical music, one was happy music (either Beethoven's Symphony #9: Presto (1min) plus Tchaikovsky's 1812 Overture (excerpt) (2mins) or Tchaikovsky's The Nutcracker: Dance of the Flutes (2mins) plus Tchaikovsky's The Nutcracker: Trepak (1min)), and the other was sad music (either Albinoni's Adagio in G Minor (3min) or Stravinsky's Firebird: Lullaby (excerpt) (3mins)). According to [19], the selected happy classical music will induce happy moods in subjects and the selected sad classical music will induce sad moods.

3.2 Design and Results of the Experiment

Warm up: Six different facial expressions of the emotional robot CIM were shown to all subjects.

Experiment: Subjects listened to a piece of classical music while simultaneously being shown a 3-minute sequence of facial expressions (either positive or negative) of the emotional robot.

Responses: After the robot emotions, subjects were asked to answer the following questions:

1. As a total impression, please select what kind of emotional music you think you were listening to from the following given choices.

A: Happy Music B: Neutral Music C: Sad Music

2. As a total impression, please select what kind of emotion (affect) you think the robot was feeling from the following given choices

A: Positive Affect B: Neutral Affect C: Negative Affect

Subjects were divided into four groups according to different combinations of classical music (happy vs. sad) and robot expressions (positive vs. negative). The different groupings constituted the Congruent condition (group 1 happy music, positive robot and group 4 sad music, negative robot together), and the Conflicting condition (group 2 sad music, positive robot and group 3 happy music, negative robot together).

The 60 subjects (29 male and 31 female) with average age 22.63 who participated in this experiment, had various nationalities. There were 15 subjects in each group, which meant that there were 30 subjects in the Congruent Condition (group 1 and group 4) and 30 subjects in the Incongruent Condition (group 2 and group 3).

The responses to the questionnaire administered after viewing the robot were analysed. It was found that subjects' identification of the emotion conveyed by the music was affected by whether or not the robot face displayed matching emotions (see Table 1). A response was considered to be correct when the music was said to have been happy in the Happy Music Condition, or sad in the Sad Music Condition (neutral music was not counted in both conditions). The Happy music was correctly recognised as such 100% of the time when it was accompanied by a "happy" robot, and only 73% of the time when accompanied by a "sad" robot. The Sad music was recognised as such 93% of the time when accompanied by a "sad" robot, and only 53% of the time when accompanied by a "happy" robot.

Table 1. Subjects' perception of the classical music

%match		Happy Music	Neutral Music	Sad Music	% correct
Happy Music Condition	group 1 (congruent robot)	100	0	0	100
	group 3 (conflicting robot)	73.3	26.7	0	73.3
Sad Music Condition	Group 2 (conflicting robot)	26.7	20	53.3	53.3
	group 4 (congruent robot)	0	6.7	93.3	93.3

Table 2 shows that subjects' judgements about the robot were also affected by the accompanying music. A response was considered to be correct when the robot head was said to have shown Positive Affect in the Positive Affect Condition, or Negative Affect in the Negative Affect Condition (the neutral choice was counted as wrong in both conditions). When the robot's positive expressions were accompanied by "happy" music, they were correctly recognised as such 100% of the time, as

Table 2. Subjects' perception of the facial expressions of the robot

%match		Positive Affect	Neutral Affect	Negative Affect	% correct
Positive Affect Condition	group 1 (congruent music)	100	0	0	100
	group 2 (conflicting music)	60	6.7	33.3	60
Negative Affect Condition	group 3 (conflicting music)	53.3	6.7	40	40
	group 4 (congruent music)	6.7	6.6	86.7	86.7

compared to only 60% of the time when accompanied by "sad" music. Conversely, the robot's expressions were correctly recognised as sad 87% of the time when paired with "sad" music, as opposed to 40% of the time when paired with "happy" music.

A Chi-square test for independence (with Yates Continuity Correction) indicated significant association between Information Style (Conflicting Information or Congruent Information) and Accuracy of recognizing robot's emotions. In other words, there was a significant difference in the accuracy of the subjects' perception of the robot's emotions depending on whether the robot's expressions and the classical music showed conflicting emotions (a medium accuracy, 15/30 (50% correct, 50% incorrect)), or congruent emotions (a relatively higher accuracy, 28/30 (93.3% correct, 6.7 % incorrect)). χ^2 (1, n=60) =11.819, p=0.001, correlation coefficient phi=-0.481 (medium effect). Consequently, H1 that when there is a surrounding musical context, people will be better at recognizing robot emotions when that context is congruent with the emotional valence of the robot emotions than when the context is incongruent with the emotional valence of the robot emotions, was supported.

Since H1 was supported in the above case, H2 that when subjects are presented with conflicting information from the robot's face and an accompanying emotional context at the same time, their perception of the robot emotions will be more influenced by the surrounding context than the robot face itself, could also be tested. A Fisher's exact test indicated the classical music did not have a more dominant effect on the judgments about the emotions of the robot than the robot emotions themselves (19/30 accuracy for classical music VS. 15/30 accuracy for Robot Emotions with two-tailed P value equals 0.4348). Consequently, H2 was not supported in this case.

Interestingly, though not one of our hypotheses, it was found that the synthetic robot emotions reversely could colour subjects' perception of the classical music. A Chi-square test for independence (with Yates Continuity Correction) indicated significant association between Information Style (Conflicting Information or Congruent Information) and Accuracy of recognizing the classical music. In other words, there was a significant difference in the accuracy of the subjects' perception of the classical music depending on whether the classical music and the robot's expressions showed conflicting emotions (a medium accuracy, 19/30 (63.3% correct, 36.7% incorrect)), or congruent emotions (a relatively higher accuracy, 29/30 (96.7% correct, 3.3 % incorrect)). χ^2 (1, n=60) =8.438, p=0.004, correlation coefficient phi=-0.417 (medium effect).

4 Discussion

This study has examined how a surrounding context (congruent or incongruent classical music) influenced users' perception of a robot's simulated emotional expressions. Hypothesis 1, that when there is a surrounding musical context, people will be better at recognizing robot emotions when that context is congruent with the emotional valence of the robot emotions than when the context is incongruent with the emotional valence of the robot emotions, was supported. However, the second hypothesis, that when a robot's expressions are not appropriate given the context, subjects' judgments of those expressions will be more affected by the context than the expressions themselves was not validated. It was found that the contextual influence was bidirectional: the music influenced judgments of the robot's emotional expressions, and the robot's expressions affected people's judgments of the emotional valence of the music.

The results reported here cohere with the limited situational dominance account [10], although that account was proposed in the context of the recognition and interpretation of *human* facial expressions. The limited situational dominance account views the facial expression itself as only one element in the interpretation of emotional expressions – others being the surrounding situation, and also the current state of the observer. The present study extends our knowledge about the kinds of context that affect the recognition of robot emotional expressions. It seems that emotionally valenced music can be shown to affect such recognition, as well as recorded speech with an emotional content, and affective pictures.

However, although the present study provides evidence of another form of context that affects the recognition of robot emotional expressions, questions remain. Would a contextual effect still be found if the context preceded, or followed the sequence of robot expressions? Will similar effects be found for any kind of context? The present results indicate that music has a different influence than the surrounding situations previously investigated. The results obtained here indicated a bidirectional effect of a musical context, which is consistent with earlier studies [11, 12, 14] with avatars or human beings. However, in our earlier study [17] we found that other emotional contexts had a more dominant effect over the robot face – subjects were more affected by the emotional content of BBC news, and by the content of affective pictures in their judgments of the robot's emotions than they were by the content of the expressions themselves. Some other researchers also found a stronger effect of surrounding context on the recognition of avatar and human faces [10, 13, 15, 16]. It seems that we do not yet have a full understanding of the factors that determine the relative effects of different forms of context.

Further research is needed, for the issue of the relationship between emotions and the surrounding situation is a complex one. What this research does indicate is that it is important to look at the relationship between a robot's expressions and what is going on around it. Previous investigations have tended to assess the extent to which robots' expressions can be recognized in neutral contexts. The robot head used here has an admittedly limited ability to articulate expressions. It lacks skin and hair and has limited degrees of freedom: left eye and right eye – yaw and tilt abilities; left eyebrow and right eyebrow – capable of raising the right and left sides; upper and lower lip – capable of bending up and down; neck – allowing the head to turn and tilt; nose – allowing movement backward and forward; cheeks – capable of rotating up and down. It might be that contextual effects were found here precisely because the robot's expressions were somewhat ambiguous. It would be interesting to know whether some robots, such as Kismet, or Probo, with more expressive faces are similarly susceptible to the effects of a surrounding context or not. However, given that even judgments about human facial expressions have been shown to be influenced by surrounding contexts, it seems likely that such robots would be affected, at least to some degree. Research by Becker-Asano and Ishiguro [20] compared recognition rates for emotional expressions displayed on Geminoid F's face, and on the face of the human model for Geminoid F. Even though there was a very close match between the robot and the human model face, higher rates of recognition were found for the human face. This suggests that robot facial expressions are also likely to be more ambiguous than human facial expressions.

In summary, previous research had shown that the recognition of human and avatar emotions can be affected by a surrounding emotional context, and the extent to which it matches, or conflicts with those emotions. The present study, together with [17], provides evidence that the recognition of the emotional expressions of a moving robot head can also be affected by a surrounding context. The present study shows that emotionally valenced music, in particular, can exert such an effect (although its effects are bidirectional in contrast to previously investigated pictorial and spoken contexts). In general, these findings do have significant implications for the designers of social and emotional robots. It seems that the match between a robot's simulated emotions and the surrounding context is important and that the interpretation of a robot's emotional expressions can be affected by what is happening around them. Hong et al. [12] used neural networks to map real-time surrounding emotional contexts to synthetic avatar emotions – similar approaches in which the valence of surrounding context is detected, and the robot's expressions adapted to match it, may well prove to be a good way of creating robots in the future that are seen as more convincing and believable.

References

1. Goris, K., Saldien, J., Vanderniepen, I., Lefeber, D.: The Huggable Robot Probo, a Multi-disciplinary Research Platform. In: Proceedings of the EUROBOT Conference 2008, Heidelberg, Germany, pp. 63–68 (2008)
2. Breazeal, C.L.: Designing Sociable Robots. In: A Bradford Book. The MIT Press (2002)
3. Ekman, P., Friesen, W.: Facial Action Coding System. Consulting Psychologists Press (1978)
4. Ekman, P., Friesen, W., Hager, J.: Facial Action Coding System, Research Nexus, Salt Lake City, Utah (2002)

5. Russell, J.: Reading emotions from and into faces: resurrecting a dimensional–contextual perspective. In: Russell, J., Fernandez-Dols, J. (eds.) The Psychology of Facial Expression, pp. 295–320. Cambridge University Press, Cambridge (1997)
6. Posner, J., Russell, J., Peterson, B.: The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology. Dev Psychopathol 17(03), 715–734 (2005)
7. Niedenthal, P.M., Kruth-Gruber, S., Ric, F.: What Information Determines the Recognition of Emotion? In: The Psychology of Emotion: Interpersonal Experiential, and Cognitive Approaches. Principles of Social Psychology series, pp. 136–144. Psychology Press, New York (2006)
8. Izard, C.E.: The face of emotion. Appleton-Century-Crofts, New York (1971)
9. Ekman, P.: Universals and cultural differences in facial expressionsof emotion. In: Cole, J.K. (ed.) Nebraska Symposium on Motivation, vol. 19, pp. 207–283. University of Nebraska Press, Lincoln (1972)
10. Carroll, J.M., Russell, J.A.: Do facial expressions signal specific emotions? Judging emotion from the face in context. Journal of Personality and Social Psychology 70, 205–218 (1996)
11. de Gelder, B., Vroomen, J.: The perception of emotions by ear and by eye. Cognition & Emotion 14, 289–311 (2000)
12. Hong, P., Wen, Z., Huang, T.: Real-time speech driven expressive synthetic talking faces using neural networks. In: IEEE Transaction on Neural Networks (April 2002)
13. Noël, S., Dumoulin, S., Lindgaard, G.: Interpreting Human and Avatar Facial Expressions. In: Gross, T., Gulliksen, J., Kotzé, P., Oestreicher, L., Palanque, P., Prates, R.O., Winckler, M. (eds.) INTERACT 2009. LNCS, vol. 5726, pp. 98–110. Springer, Heidelberg (2009)
14. Creed, C., Beale, R.: Psychological Responses to Simulated Displays of Mismatched Emotional Expressions. Interacting with Computers 20(2), 225–239 (2008)
15. Mower, E., Lee, S., Matariric, M.J., Narayanan, S.: Human perception of synthetic character emotions in the presence of conflicting and congruent vocal and facial expressions. In: IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP 2008), Las Vegas, NV, pp. 2201–2204 (2008)
16. Mower, E., Matarić, M.J., Narayanan, S.: Human perception of audio-visual synthetic character emotion expression in the presence of ambiguous and conflicting information. IEEE Transactions on Multimedia 11(5) (2009)
17. Zhang, J., Sharkey, A.J.C.: Contextual Recognition of Robot Emotions. In: Groß, R., Alboul, L., Melhuish, C., Witkowski, M., Prescott, T.J., Penders, J. (eds.) TAROS 2011. LNCS (LNAI), vol. 6856, pp. 78–89. Springer, Heidelberg (2011)
18. Bradley, M.M., Lang, P.J.: The International Affective Picture System (IAPS) in the Study of Emotion and Attention. In: Coan, J.A., Allen, J.B. (eds.) The Handbook of Emotion Elicitation and Assessment, pp. 29–46 (2007)
19. Eich, E., Ng, J.T.W., Macaulay, D., Percy, A.D., Grebneva, I.: Combining music with thought to change mood. In: Coan, J.A., Allen, J.B. (eds.) The Handbook of Emotion Elicitation and Assessment, pp. 124–136. Oxford University Press, New York (2007)
20. Becker-Asano, C., Ishiguro, H.: Evaluating facial displays of emotion for the android robot Geminoid F. In: IEEE SSCI Workshop on Affective Computational Intelligence, pp. 22–29 (2011)

Analysis of Bluffing Behavior in Human-Humanoid Poker Game

Min-Gyu Kim[1] and Kenji Suzuki[1,2]

[1] Dept. of Intelligent Interaction Technologies, University of Tsukuba, 1-1-1, Tennodai, Tsukuba, Japan
[2] PRESTO, Japan Science and Technology Agency
mingyu@ai.iit.tsukuba.ac.jp,
kenji@ieee.org

Abstract. This paper presents the analysis of human nonverbal responses and betting decision in terms of bluffing and the comparison between human-human poker game and human-robot poker game. According to a given situation manipulated with the card hand strength, we explored how different the participants change bluff decision between strong hand and weak hand and how different the participants use a bluff between a human-human and human-robot poker game. Furthermore, we analyzed the significant correlation between the card hand strength and human behaviors and obtained the regression model to predict the card hand strength from nonverbal behaviors and bluff decision.

Keywords: Human-robot social interaction, Humanoid playmate, Poker game, Bluff analysis.

1 Introduction

Poker game is one of social activities that intensively show human complex behaviors such as strategizing, bargaining and bluffing. In an ordinary poker game, people engage in social interactions while they play attentively to gain advantage. Since poker game needs not only logical thinking but also psychological skills such as a poker face, it is strongly required to build a poker playing robots being able to understand human social interactions.

Many empirical studies have emphasized the importance of the role of physical embodiment in human-robot interaction. In [1], Wainer *et al* measured differences in perception of social interaction comparing with virtual agents and remote physical robots. In [2], Lee *et al* evaluated the effect of physical embodiment on the feeling of an agent's social presence, and showed that lonely people provide more positive social responses to social agents than non-lonely people. [3] reported how the robot's physical presence influences a person's perception in multidimensional characteristics and the differences between a robot and an animated character in terms of engagement and perceptions. In [4], Kanda *et al* have presented measured human body movements while the human observed and interacted with the physical robot.

B. Mutlu et al. (Eds.): ICSR 2011, LNAI 7072, pp. 183–192, 2011.

In this study, we focused on human behavioral responses in the environment where people communicate with an interactive and engaging humanoid robot in a particular social context. We considered that a poker game is a good example of human-robot social interaction. We explored the possibilities of the poker playing humanoid not only being able to understand the human behaviors but also perform complex behaviors due to human behavior in terms of deception. Social psychological studies about impression management have presented that people in various social situations use the displayed aspect of the self for strategic purpose [5,6]. A typical deceptive behavior in poker game is bluffing which is a useful action for game strategy to hide evidences of own card hand. Poker players can use bluff in betting decisions or by giving false nonverbal information to opponents whereas they can be affected by opponent's strategic bluff as well. In this sense, it is important to design a poker-playing robot that executes behaviors determined by human betting decision and nonverbal behaviors as well as to build robots that influence people in poker game. However, since it is not easy to analyze human social interaction in poker game, studies about a humanoid robot that is able to perform human-like psychological behaviors such as bluffing, lying, etc have not been sufficiently researched yet. Therefore, we examined the possibilities of developing robot's decision-making structure based on human nonverbal behaviors and betting decision.

In Section 2, the experimental setup, procedure and conditions are described. In Section 3 and 4, we discussed the analysis results between human-human poker game and human-robot poker game.

2 Experiment

2.1 Humanoid Robot Platform

The autonomous poker-playing humanoid has been built to investigate human behavioral analysis [7]. The developed humanoid robot is composed of the upper

(a) Humanoid playmate (b) Human-robot poker game (left) and human-human poker game (right)

Fig. 1. Experimental setup

torso with a wall-mounted waist as depicted in Figure 1(a). It has totally 32DOF in the body (8DOF for the head, 3DOF for the waist, 7DOF for the right arm, 4DOF for the right hand, 5DOF for the left arm, and 5DOF for the left hand).

The humanoid robot system is divided into vision, audio and robot control software modules. The vision module includes facial feature detection, hand movement tracking, eye blink detection and card recognition. The auditory module records the conversation during the game and synthesizes speech for betting. The robot control module performs the position control of all the joint angles of the humanoid. However, the automatic detection of human nonverbal behaviors was not used because we focused on behavior analysis in this study.

2.2 Procedure

We recruited ten subjects (age 22-29, 6 males and 4 females) and a paid dealer and then obtained the subject's informed consent. The subjects were not professional players and some of them have never played poker before. Before the experiment, we trained the dealer to master the rule of Texas hold'em and the purpose of the experiment. Also the dealer was carefully instructed not to talk and smile to the subjects because the dealer's speech and behaviors might affect the subjects' behaviors. The dealer could speak only to notify the betting turn to the players.

Each subject played Texas hold'em first with the experimenter and then with the humanoid one on one as shown in Figure 1(b). In a human-robot poker game, the experimenter controlled the humanoid's movements behind the experimental space. Based on Wizard of Oz approach [8], the humanoid randomly performed affirmative head nods, negative head shakes and head orientation. Before beginning the experiment, the dealer explained the rule of Texas hold'em to the subjects and then they practiced the game until they got a hold of the game. The subjects started to play Texas hold'em with fake money. In this study, in order to save the playing time, the dealer assisted the humanoid in handling the cards and the fake money. Instead, the humanoid raised the stakes or folded by saying "bet one thousand yen" or "fold". All the subjects played five rounds of the poker game. The entire episode was recorded using a video camera installed in right side of the experimenter in a human-human poker game, and the video cameras in a human-robot poker game installed in the humanoid's body and right side of the humanoid.

2.3 Simplified Rule of Texas Hold'em

Each betting round, the participants can bet from a thousand yen to three thousand yen and used play bills. The players also included a blind bet before the card distribution. As usual, the players share five community cards that are opened on the table. After the blind bet, the players place their bet with the first three community cards (flop cards) and their cards (pocket cards). When the tokens which a player bet would be equally matched by his opponent, the first betting round ends and the fourth community card (turn card) is opened. Then the players begin to bet money again. The round closes when the bet is matched equally by the opponent. The final community card (river card) is opened and final bet is placed. After the end of the final betting round, the players open their pocket cards for a showdown. If one of the players folds in the middle of the game, the game is over and the pocket cards are not revealed.

2.4 Card Hand Condition

A poker player's nonverbal behaviors are related to his own card hand [9]. In order to see the effect of card hand strength on human response, the poker game of five rounds was manipulated prior to the experiment. Given card hands were divided into strong hand and weak hand conditions. The participants played Texas hold'em with the strong card hands in 2nd and 5th rounds and with the weak card hands in 1st, 3rd and 4th rounds. The strong card hands were straight and four of a kind, which means five consecutive cards and four cards of the same rank, respectively. The weak card hands indicated no pair to make them lose.

2.5 Measurement

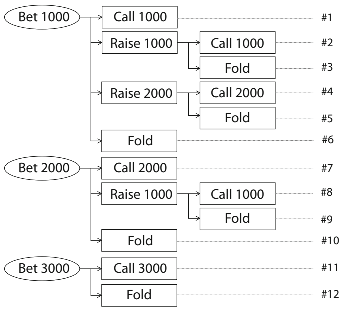
Bluff. In a poker game, bluff strategy is usually dependent on the card hand strength. Therefore, a poker player with a winning hand would try to bet carefully to keep the pot growing and at the same time keep the opponent from folding early. On the other hand, the participant with a losing hand would try to bet in a way that the other players would assume otherwise and raise the bet taking high risks. Based on these assumptions, the player with a strong hand would avoid maximum betting and the player with a weak hand would avoid minimum betting, is considered as the bluff strategies.

Nonverbal behaviors. Smile [10,11], hand movement [12], eye blink [13] and eye gaze [14] were chosen because they have been widely discussed in a high stake, deceptive situation which is also inherent in poker game.

According to [14], the smile was defined as smiling as perceived by the coders, who were given no specific definition or were given a definition not involving specific AUs. We coded the amount of smile from recorded video and then calculated frequency of smile that is a total amount of smile divided by a minute.

In a deceptive situation, people usually show two kinds of hand movement, adaptor and illustrator [14]. Adaptors are movements indicating a low level of awareness such as self-touching and illustrator is a gesture performed while talking, illustrating and assisting the speech. We considered that because our participants would not talk so much, the self-touching would be mostly detected in a poker game. The self touching was classified into three types of hand movement, which are 'touching face', 'touching head' and 'touching arm'. We collected the amount of each hand movement and calculated frequencies of touching face, touching head, and touching arm, respectively.

Eye blink was defined as eyes opening and closing quickly, and eye gaze was defined as facing the other person/objects and gazing at the person/objects [14].



(a) Betting decision tree

Outcome	Strong hand		Weak hand	
No.	Subject	Opponent	Subject	Opponent
1	Bluff	Bluff	Truth	Truth
2	Truth	Bluff	Bluff	Bluff
3	Bluff	Bluff	Truth	Bluff
4	Truth	Truth	Bluff	Bluff
5	Bluff	Truth	Truth	Bluff
6	Bluff	Bluff	Truth	Truth
7	Bluff	Bluff	Bluff	Bluff
8	Truth	Truth	Bluff	Bluff
9	Bluff	Truth	Truth	Bluff
10	Bluff	Bluff	Bluff	Truth
11	Truth	Truth	Bluff	Bluff
12	Truth	Bluff	Bluff	Truth

(b) Result table for bluff and truth

Fig. 2. Judgement of bluff based on betting decision tree

Especially, the eye gaze was divided into three groups; ‘gazing at opponent’, ‘gazing at table’ and ‘gazing at other’. We counted the amount of eye blink and eye gaze and then transformed them to frequencies of eye blink and eye gaze, respectively.

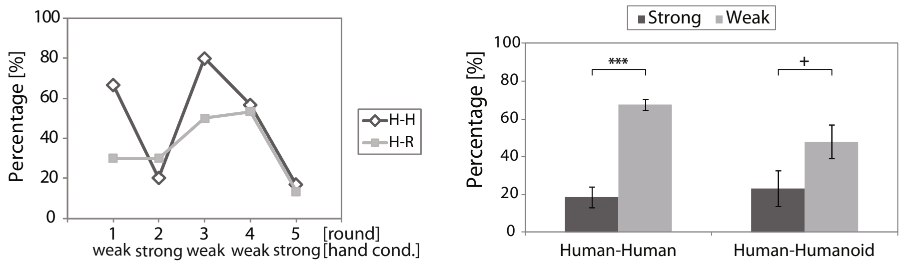
Descriptive questionnaire. The two subjective type questions were given to the participants to describe how they felt while robot moved its head and arm to play game and how they thought about the difference between human and robot opponents.

3 Results

3.1 Bluff in Betting Decision

All possible cases with betting order in three columns and possible subsequent choices with respect to opponent’s decision in twelve rows are illustrated in Figure 2(a). For instance, if the opponent bets 1000 yen in the first betting order, the participant can make one of given four decisions (i.e. call 1000 yen, raise 1000 yen, raise 2000 yen or fold) in the second betting order. Since all betting decisions of the participants belonged in the betting decision tree without exception, we could determine the subjects’ bluff decisions based on card hand strength and first player with betting priority as shown in Figure 2(b).

The average percentage of bluff over game rounds was assessed using the Friedman non-parametric test as shown in Figure 3(a). It was assumed that the card hand strength effects on human behaviors if there would be significant difference between 1st-2nd (weak-strong), 2nd-3rd (strong-weak) and 4th-5th rounds (weak-strong) in which the card hand strengths are completely different from each other. In human-human poker game, the card hand strength significantly influenced on the bluff strategy of subjects ($\chi^2(4)$=18.14, p<.05). According to



(a) Change of bluff over game rounds

(b) Avg. percentage of bluff with standard error intervals (Note: $^{***}p<.01$, $^{+}p<.10$)

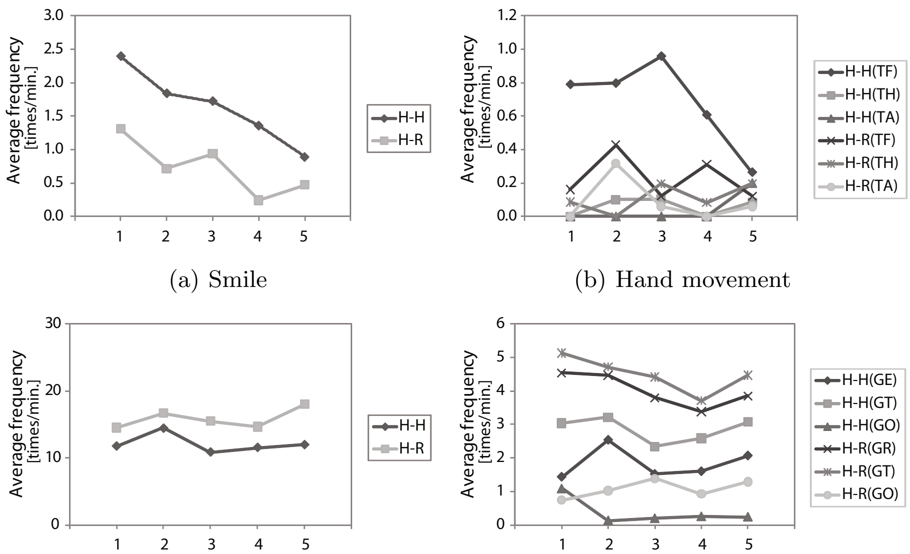
Fig. 3. Bluff analysis

a pair-wise comparison using Wilcox test with FDR correction, there was a significant difference in average percentage of bluff between 1st-2nd, 2nd-3rd and 4th-5th rounds ($p<.05$, respectively). The bluff decisions seemed to be affected by the card hand strength accordingly. In human-robot poker game, there was no significance in average percentage of bluff.

3.2 Effect of Card Hand Strength on a Nonverbal Behavior

The Kruskal-Wallis test was performed to examine how the participants made bluff decisions according to the card hand strength in the overall game rounds as shown in Figure 3(b). In the human-human poker game, the average percentage of bluff changed from 18.3% (strong card hand) to 67.8% (weak card hand), showing a highly significant difference between strong hand and weak hand ($\chi^2(1)=14.65$, $p<.01$). On the other hand, in the human-robot poker game, the average percentage of bluff was increased from 21.6% (strong card hand) to 47.8% (weak card hand), reflecting marginally significant difference between strong hand and weak hand ($\chi^2(1)=3.8$, $p<.10$).

As can be seen from Figure 4, the frequency variations of four nonverbal behaviors over game rounds were investigated using the Friedman test. As the game rounds progressed, the frequency of smile tended to be significantly decreased (H-H: $\chi^2(4)=9.78$, $p<.05$, H-R: $\chi^2(4)=13.97$, $p<.05$). However, in the pair-wise comparison using Wilcox test with FDR correction, there were no significant differences between 1st-2nd, 2nd-3rd and 4th-5th rounds. The result does not imply that the smile frequency of participants is dependent on the card hand strength (See Figure 4(a)). The frequency variations of hand movement such as Touching Face (TF), Touching Head (TH) and Touching Arm (TA) showed no significance in the effect of card hand strength as depicted in Figure 4(b). As shown in Figure 4(c), in case of eye blink, there were marginally significant differences (H-H: $\chi^2(4)=8.06$, $p<.10$, H-R: $\chi^2(4)=8.48$, $p<.10$). However, the pair-wise comparison showed no significant differences between 1st-2nd, 2nd-3rd and 4th-5th rounds. The card hand strength also was not effective on eye gaze, although there was



(a) Smile (b) Hand movement

(c) Eye blink (d) Eye gaze

Fig. 4. The frequency variation of nonverbal behaviors over the game rounds

significance (Gazing at Other (GO): $\chi^2(4)$=12.09, p<.05) in human-human poker game and significances (Gazing at Robot (GR): $\chi^2(4)$=9.52, p<.05, Gazing at Table (GT): $\chi^2(4)$=12.77, p<.05) in human-humanoid poker game (See Figure 4(d)). It is because the pair-wise comparison did not indicate significances in frequency variation of eye gaze between 1st-2nd, 2nd-3rd and 4th-5th rounds.

3.3 Regression Model of Card Hand Strength

As known in the previous section, each nonverbal behavior was not independently effective factor in estimating the card hand strength because of the extremely complicated human behaviors in the human-humanoid poker game. The Spearman non-parametric correlation test also supports this result except for the significant correlation between card hand strength and bluff decision (Spearman's rho=-0.88, p<.01). Therefore, in order to model human opponent considering decisions and social aspect of interaction for poker game altogether, we found the multiple regression model of card hand strength determined by bluff decision and the multiple human nonverbal behaviors.

The regression model between card strength and the nine factors of human behaviors was investigated using the multiple regression analysis with stepwise selection. Table 1 shows the significant independent variables, partial regression coefficient, p-value and standardized partial regression coefficient which are constitutive of the obtained regression model with a high significance.

Table 1. Regression coefficients obtained by multiple regression analysis with stepwise selection (Note: $^{***}p<.001$, $^{**}p<.01$, $^{*}p<.05$)

Independent variables	Partial regression coefficient	Standardized partial regression coefficient
Bluff (α_B)	-0.018	-1.046^{***}
Touching Face (α_{TF})	0.101	0.269^{*}
Touching Head (α_{TH})	1.141	0.368^{*}
Touching Arm (α_{TA})	-1.070	-0.466^{**}
Gazing at Robot (α_{GR})	-0.102	-0.626
Gazing at Table (α_{GT})	0.094	0.502
Gazing at Other (α_{GO})	0.102	0.269^{*}
Intercept	0.983	

Equation (1) is the derived regression model of card hand strength. The analysis resulted in seven significant human behaviors that contribute to the estimation of card strength, except for smile and eye blink. The self touching (Touching Face: R^2=0.27, $p<.05$, Touching Head: R^2=0.37, $p<.05$, and Touching Arm: R^2=-0.47, $p<.01$) and the eye gaze (Gazing at Robot: R^2=-0.62, Gazing at Table: R^2=0.50, Gazing at Other: R^2=0.27, $p<.05$) were correlated with the regression model of card hand strength. The multiple correlation coefficient is 0.89, and 80.5% of the variability can be accounted for by analysis of variance (F(7,12)=14.96, $p<.001$).

$$CS = 0.983 + \alpha_B B + \alpha_{TF} TF + \alpha_{TH} TH + \alpha_{TA} TA + \alpha_{GR} GR + \alpha_{GT} GT + \alpha_{GO} GO \quad (1)$$

4 Discussion

The experiment results revealed two important points that offer the effects of given situations on human behaviors and also the possibilities of an interactive poker playing humanoid in terms of deception. While participants played a poker game with the humanoid, they showed very complex responses through bluff decision and nonverbal behaviors.

The smile of participant was not influenced by the card hand strength and even not a significant factor in the regression model of the card hand strength. According to the descriptive answers of questionnaire, some participants attempted to control their facial expressions because they felt that informative head movements of the humanoid sometimes created a sense of observing them. This kind of answer is fairly associated with a research finding that people manage facial expressions in deceptive situation [13]. We supposed that the management of smile is probably more related to engagement for a poker game. In other words, since the participants were engaged in the poker game with the humanoid, they seemed to accept that seriously as like as a real poker game with human where deception is naturally inherent.

The card hand strength did not have an independent effect on the touching face, head and arms, respectively. However, in conjunction with eye gaze and bluffing rate, the hand movements showed significance in the regression model of card hand strength. The frequencies of touching face and touching head were significantly related to the card hand strength, reflecting that the self touching is increased in deceptive communication [9]. The participants seemed not to manage these kinds of hand movements consciously while attempting to control the smile. On the other hand, an inverse proportion of the frequency of touching arm, mostly cross-arms, with respect to the card hand strength shows that the participants sometimes controlled their hand movements consciously by crossing arms [11].

Eye blink had no significance in an independent effect of card hand strength, and no correlation with the regression model of card hand strength. However, eye gaze, especially 'gazing at other', was significantly related to the estimation of card hand strength in this experiment. As gazing at other is in directly proportion to card hand strength, the participants with the strong card hand tended to look off to other places rather than the weak card hand. Gazing at other is a kind of gaze aversion, meaning that the participants seemed not to contact eyes with the humanoid [11].

Finally, we obtained another significant result from the analysis of bluff decision which was substantially dependent on the card hand strength. In the early stage of designing the experiment, we assumed that although the subjects are non-poker player, they would bet carefully in non-threatening way that an opponent would assume otherwise and take high risks. Against the expectation, the betting decision was affected by card hand strength except for the initial round where subjects did not try to bet deceptively. Meanwhile, we considered that if the subjects would get hold of Texas hold'em much more with technical skills, the effect of card hand strength would be accordingly decreased over time. Therefore, two requirements for a decision-making of autonomous poker-playing humanoid emerged from the results of this study. The first requirement is to perform complex behaviors due to human betting decision and nonverbal behaviors. The second requirement is to learn a given situation with evidences observed in other situations created by not only human behaviors but also effect of its behaviors.

5 Conclusion

We have found difference in human behaviors between human-human poker game and human-robot poker game, and also that bluff decisions and nonverbal responses were significantly related to predicting card hand strength of opponent. These findings provide possibility to build a humanoid robot to perform complex behaviors by modeling human opponent based on given situation and human responses in a poker game.

In the near future, a further psychological research will be conducted in order to find significant and specific robot behaviors to drive people to engage in the poker game with the developed humanoid platform. At the same time, a decision making of the humanoid playmate will be designed to be able to understand and learn human behaviors while playing the poker game autonomously.

Acknowledgments. This work is partially supported by Grant-in-Aid for Scientific Research and Global COE Program on "Cybernics: fusion of human, machine, and information systems" by MEXT, Japan.

References

1. Wainer, J., Feil-Seifer, D.J., Shell, D.A., Mataric, M.J.: The Role of Physical Embodiment in Human-Robot Interaction. The 15th IEEE Int. Symp. on Robot and Human Interactive Communication, 117–122 (2006)
2. Lee, K.M., Jung, Y., Kim, J., Kim, S.R.: Are Physically Embodied Social Agents Better than Disembodied Social Agents?: The Effects of Physical Embodiment, Tactile Interaction, and People's Loneliness in Human-Robot Interaction. Int. J. Human-Computer Studies 64(10), 962–973 (2006)
3. Bainbridge, W.A., Hart, J.W., Kim, E.S., Scassellati, B.: The Effect of Presence on Human-Robot Interaction. In: 17th IEEE Int. Sym. on Robot and Human Interactive Communication, ROMAN 2008 (2008)
4. Kanda, T., Ishiguro, H., Imai, M., Ono, T.: Body Movement Analysis of Human-Robot Interaction. Int. Joint Conf. on Artificial Intelligence 18, 177–182 (2003)
5. Baumeister, R.F.: A Self-presentational View of Social Phenomena. Psychological Bulletin 91(1), 3–26 (1982)
6. Schlenker, B.R., Pontari, B.A.: The Strategic Control of Information: Impression Management and Self-presentation in Daily Life. In: Tesser, A., Felson, R.B., Suls, J.M. (eds.) Psychological Perspectives on Self and Identity, pp. 199–232 (2000)
7. Kim, M.G., Suzuki, K.: A Card Playing Humanoid for Understanding Socio-emotional Interaction. In: Yang, H.S., Malaka, R., Hoshino, J., Han, J.H. (eds.) ICEC 2010. LNCS, vol. 6243, pp. 9–19. Springer, Heidelberg (2010)
8. Kelley, J.F.: An Iterative Design Methodology for User-Friendly Natural Language Office Information Applications. ACM Trans. on Office Information Systems 2(1), 26–41 (1984)
9. Hayano, D.M.: Communicative Competency among Poker Players. J. of Communication 30(2), 99–104 (1980)
10. Ambadar, Z., Schooler, J.W., Cohn, J.F.: Deciphering the Enigmatic Face: the Importance of Facial Dynamics in Interpreting Subtle Facial Expressions. Psychol. Sci. 16(5), 403–410 (2005)
11. Valstar, M.F., Pantic, M., Ambadar, Z., Cohn, J.F.: Spontaneous vs. Posed Facial Behavior: Automatic Analysis of Brow Actions. In: 8th Int. Conf. on Multimodal Interfaces, pp. 162–170 (2006)
12. Vrij, A., Edward, K., Roberts, K.P., Bull, R.: Detecting Deceit via Analysis of Verbal and Nonverbal Behavior. J. of Nonverbal Behavior 24(4), 239–263 (2004)
13. Fukuda, K.: Eye Blinks: New Indices for the Detection of Deception. Int. J. of Psychophysiology 40(3), 239–245 (2001)
14. DePaulo, B.M., Lindsay, J.J., Malone, B.E., Muhlenbruck, L., Carlton, K., Cooper, H.: Cues to Deception. Psychological bulletin 129(1), 74–118 (2003)

Evaluating Supportive and Instructive Robot Roles in Human-Robot Interaction

Manuel Giuliani[1] and Alois Knoll[2]

[1] fortiss GmbH, Guerickestraße 25, 80805 München
giuliani@fortiss.org
[2] Technische Universität München, Robotics and Embedded Systems,
Boltzmannstraße 3, 85748 Garching
knoll@in.tum.de

Abstract. Humans take different roles when they work together on a common task. But how do humans react to different roles of a robot in a human-robot interaction scenario? In this publication, we present a user evaluation, in which naïve participants work together with a robot on a common construction task. The robot is able to take different roles in the interaction: one group of the experiment participants worked with the robot in the *instructive role*, in which the robot first instructs the user how to proceed with the construction and then supports the user by handing over building pieces. The other group of participants used the robot in its *supportive role*, in which the robot hands over assembly pieces to the human that fit to the current progress of the assembly plan and only gives instructions when necessary. The results of the experiment show that the users do not prefer one of the two roles of the robot, but take the counterpart to the robot's role and adjust their own behaviour according to the robot's actions. This is revealed by the objective data that we collected as well as by the subjective answers of the experiment participants to a user questionnaire. The data suggests that the most influential factors for user satisfaction are the number of times the users picked up a building piece without getting an explicit instruction by the robot and the number of utterances the users made themselves. While the number of pickup actions had a positive or negative influence, depending on the role the users took, the number of own utterances always had a strong negative influence on the user's satisfaction.

1 Introduction and Related Work

When humans work together, they take different roles in the interaction. For example when two persons assemble a shelf, usually one of them takes the lead and gives instructions on how to follow the assembly plan. The other person helps the instructor to build the shelf and to gather the right parts for the next building step. The question that we are following in this publication is: in a similar construction task, do humans prefer a robot that takes the role of an instructor or a supporter?

For that, we conduct a human-robot interaction experiment in which naïve participants have to build target objects from a wooden toy construction set

B. Mutlu et al. (Eds.): ICSR 2011, LNAI 7072, pp. 193–203, 2011.

together with a robot. The robot takes either the role of an instructor or a supporter. We use the robot in both settings for a between participants experiment, in which we collect objective and subjective measurements to compare the changes in behaviour and opinion about the robot between the two experiment participant groups.

In robotics research, there are two areas in which the role of a robot is of importance: on the one hand, there are robots that have to interact with humans in various scenarios. Here, the research focuses on the different roles the robot can take in the interaction and how the human partners of the robot react to these roles. On the other hand, researchers are interested in the roles of robots in multi-robot teams. Here, robots use different roles to solve a given task more effectively.

[1] were among the first authors who realised that for a social robot that is capable to interact with a human it is of importance, which social role the robot should take in this interaction. [4] presented one of the earliest studies that researched how humans react to different robot roles. They conducted an experiment, in which a human and a robot had to work together. In the experiment, the authors varied the appearance of the robot as well as the behaviour (i.e. the role) of the robot. The results of the experiment show that humans rely more on human-like robots and feel more responsible for the task when the robot looks more machine-like. The experiment also showed that the participants felt less responsible for the task when they worked with a robot who took the role of a supervisor, which is also supported by our findings.

[9] show a socially assistive therapist robot that monitors and encourages humans in rehabilitation exercises. This robot shows either an introverted or an extroverted personality. Tapus and Maraic were able to show in an experiment that introverted patients interacted significantly longer with the introverted robot, while extroverted patients interacted longer with the extroverted robot, respectively. In contrast to our findings, it seems that in this type of interaction, humans prefer to have a partner with similar personality traits to their own.

[7] argument that for urban search & rescue robots (USAR) affective computing is important. Amongst other things, they present the theoretical basis for the implementation of social roles on a USAR, so that the robot can adapt its own behaviour on a rescue mission, corresponding to whether it is interacting with a fellow helper or a victim.

2 Human-Robot Interaction System

The experiment described in this paper makes use of a completely autonomous human-robot interaction system (**Figure 1**) which supports multimodal human-robot collaboration on a joint construction task. The participant and the robot work together to assemble wooden construction toys on a common workspace, coordinating their actions through speech and gestures. The robot can pick up and move objects in the workspace and perform simple assembly tasks. In the scenario considered here, human and robot both know the assembly plan and jointly execute it. The robot assists the humans by explaining necessary assembly steps when the humans do not execute them by themselves and by offering pieces

Fig. 1. Human-robot interaction system. The robot has a pair of manipulator arms with grippers, mounted in position to resemble human arms, and an animatronic talking head [10] capable of producing facial expressions, rigid head motion, and lip-synchronised synthesised speech.

as required. The workspace is divided into two areas—one belonging to the robot and one to the human—to make joint action necessary for task success.

In this experiment, the robot shows two different roles: in the *instructive role*, the robot first gives instructions to the user how to assemble pieces according to the assembly plan before handing over construction pieces from its own work area. In the *supportive role*, the robot first hands over construction pieces from its own workspace to the human and only gives instructions if the users do not pick up the right construction pieces from their workspace.

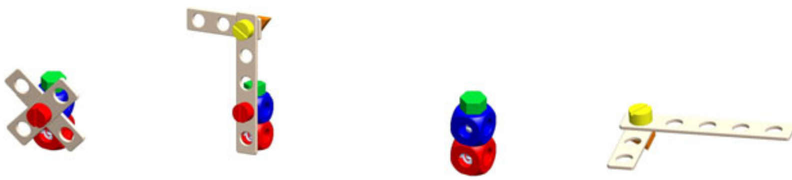
Although the construction pieces can be screwed and stuck together, the robot is not able to perform assembly actions itself. However, the robot supports its human co-worker with the following list of actions: *give*, the robot hands over a construction piece from its own workspace to the human. *tellAbout*, the robot instructs the human to pick up a piece from the human's workspace because it fits to a building step of the currently loaded plan. *askFor*, the robot asks the human to put a certain construction piece on the workspace, because it is needed for a given plan and the robot cannot detect it with its object recognition. *tellBuild*, the robot asks the human to build one of the substeps of the currently loaded plan. *thankFor*, the robot thanks the human for a construction piece that the human has put on the table.

3 Experiment

In this section, we describe the experiment set-up, demography of the experiment participants, data collection and analysis, and results.

3.1 Experiment Design

This study used a between participants design with one independent variable: each participant interacted either with the robot that used the supportive role setting, or else with a system that used the instructive role. The robot was completely autonomous and did not get any other outside information than from its speech and object recognition sensors. Each participant built two target objects in collaboration with the system, always in the same order, first the *windmill* (**Figure 2a**), after that the *railway signal* (**Figure 2b**). For both target objects, the user was given an assembly plan on paper.



(a) Windmill **(b)** Railway signal **(c)** Tower **(d)** L shape

Fig. 2. Target objects for the experiment. Both target objects consist of a base that is named *tower*. A windmill is a tower combined with two small slats; a railway signal is a tower combined with an *l shape*.

The participants stood in front of the table facing the robot, equipped with a headset microphone for speech recognition. Participants got instructions that they could speak with the robot by using a set of predefined phrases: they could either ask the robot for one of the pieces in the robot's workspace by giving a direct order, for example by saying "give me a blue cube", or they could ask the robot to repeat its last utterance by saying "pardon me?".

The pieces required for the target object were placed on the table, using the same layout for every participant. The layout was chosen to ensure that there would be enough similar construction pieces on both sides of the table for every subplan of the target objects so that the robot could either perform the action *give* and handover an object from its side of the table or the action *tellAbout* and instruct the users to pick up an object from their side of the table. For example, for the tower of the windmill there was a red cube in both table areas, so that the robot could either hand over the cube from its own workspace or instruct the participants to pick up the cube from their workspace. Along with the assembly plan mentioned above, the participants were given a table with the names of the pieces they could build the objects with.

3.2 Participants

40 participants (27 male), who were naïve in the sense that they never worked with the robot before, took part in this experiment. The mean age of the participants was 27.20 (7.06), with a minimum of 17 and a maximum of 59. Of the

participants who indicated an area of study, the two most common areas were Mathematics (11 participants) and Informatics (8 participants). On a scale of 1 ("I do not agree at all") to 5 ("I completely agree"), participants gave a mean assessment of their knowledge of computers at 3.68 (1.00), of speech recognition systems at 1.90 (1.03), and of human-robot interaction systems at 1.60 (1.01). For their participation in the experiment, the participants got the chance to win a voucher for an online shop.

3.3 Hypotheses

In this study, we compare how humans react to the instructive and supportive role of the robot when they build target objects with it. We are mainly interested if the experiment participants accept both roles of the robot or if there is a clear preference for one of the two roles. In particular, we have the following two hypotheses:

H1. Experiment participants who work with the robot in the supportive role, generally asses their interaction with the robot more positive.

H2. Experiment participants who work with the robot in the supportive role display a more proactive behaviour, while participants using the instructive robot will take a more passive role in the interaction.

Since we gathered a wide range of subjective and objective measures in this study, we did not make specific predictions as to which specific measure the experimental manipulations will have an effect.

3.4 Data Acquisition

At the end of a trial, the participants responded to a usability questionnaire consisting of 29 items, which fell into four main categories: *feelings of the user* (10 items), *intelligence of the robot* (7 items), *robot behaviour* (6 items), and *task success* (6 items). The items on the questionnaire were based on those used in two previous user evaluations [2] [3], but were adapted for the scenario and research questions of the current study. The questionnaire was presented using software that let the participants choose values between 1 ("I do not agree at all") and 100 ("I completely agree") with a slider.

In addition to the questionnaire, we collected a set of objective measurements from the automatically generated system log files and from annotations of the videos we took during the experiments. All in all, we had four different objective measurements:

- the *number of verbal utterances* by the participants, which is the number of times the users asked the robot for a certain construction piece or to repeat its last utterance,
- the *number of times the participants picked up a construction piece* from their side of the table, where the robot did not instruct them to pick up the object,

- the *number of instructions the robot gave to the participants*, i.e. the instructions in which the robot told the human which piece to pick up next from the workspace, and
- the *overall duration* the participants needed to build windmill and railway signal.

We took the first two measurements from the system log files; we annotated the videos of the experiment participants with Anvil [5] to collect the remaining two measurements. Not all participants agreed that we videotaped them, thus we only have video data for 32 of the 40 participants, 17 videos of participants who used the instructive robot and 15 videos of participants who used the supportive robot.

3.5 Results

In this study, we analysed the collected data in several ways. First, we compared the subjective answers of the experiment participants to the user questionnaire to find out if there are any significant differences between the answers of the group that worked with the supportive robot and the group that worked with the instructive robot. Second, we compared the objective measurements that we took from the system logs and the videos to find differences between the two groups. Third, we calculated which of the objective measurements could potentially predict the subjective answers by the experiment participants.

Subjective Measurements. We applied a Mann-Whitney test on the answers to the user questionnaire to analyse if the different robot roles had a significant effect on the ratings by the two participant groups. Generally, participants gave a positive feedback of an average 82.84 (20.26) on the questions of the *feelings of the user* category, in which they had to rate if their interaction with the robot was enjoyable. However, the participants rated the robot's intelligence with only 56.35 (26.16) points. There was no significant difference in these questions between the two groups.

We found significant differences (p-value < 0.05) in the ratings for 4 of the 29 statements of the user questionnaire, which are displayed in **Table 1**.

Table 1. Statements with significant differences between user groups of user questionnaire

Statement	Supportive	Instructive	M-W
I found the robot easy to use.	83.80 (12.81)	**90.80 (13.03)**	$p \approx 0.043$
I knew what I could say or do at each point in the conversation.	71.05 (30.32)	**90.10 (12.04)**	$p \approx 0.038$
It was clear what to do when the robot did not understand me.	**70.65 (21.46)**	57.33 (15.26)	$p \approx 0.034$
The robot gave too many instructions.	33.95 (28.21)	**16.71 (21.95)**	$p \approx 0.026$

Objective Measurements. We show the results of the objective measurements in **Table 2**. We computed if there is a significant difference between the two user groups, again via a Mann-Whitney test. We found a significant difference for the number of robot instructions, which is not surprising, but shows that the instructive robot gave significantly more instructions to the user. Furthermore, users who worked with the supportive robot significantly picked up more construction pieces without getting instructions from the robot to do so.

Table 2. Objective results

Measure	Instructive	Supportive	M-W
No. of user utterances	1.65 (1.69)	1.25 (1.94)	$p \approx 0.33$
No. of user actions	0.76 (0.90)	**4.80 (1.97)**	$p < 0.01$
No. of robot instructions	**10.3 (1.49)**	4.60 (2.28)	$p < 0.01$
Assembly duration (seconds)	265.86 (46.22)	258.80 (51.32)	$p \approx 0.82$

Predictive Measurements. To complete the result analysis of this study, we calculated a predictor function to compute if the objective measurements we collected in this evaluation could predict the subjective statements of the user questionnaire. Being able to predict subjective user satisfaction from more easily-measured objective properties can be very useful for developers of interactive systems: in addition to making it possible to evaluate systems based on automatically available data without the need for extensive experiments with users, such a performance function can also be used in an online, incremental manner to adapt system behaviour to avoid entering a state that is likely to reduce user satisfaction, or can be used as a reward function in a reinforcement-learning scenario [11].

To compute the predictor function, we employed a procedure similar to that used in the PARADISE evaluation framework (PAradigm for DIalogue System Evaluation) [11]. The PARADISE model uses stepwise multiple linear regression to predict subjective user satisfaction based on measures representing the performance dimensions of task success, dialogue quality, and dialogue efficiency, resulting in a predictor function of the following form:

$$Satisfaction = \sum_{i=1}^{n} w_i * \mathcal{N}(m_i)$$

The m_i terms represent the value of each measure, while the $\mathcal{N}$ function transforms each measure into a normal distribution using z-score normalisation. Stepwise linear regression produces coefficients (w_i) describing the relative contribution of each predictor to the user satisfaction. If a predictor does not contribute significantly, its w_i value is zero after the stepwise process. **Table 3** shows the predictor functions that we calculated using stepwise multiple linear regression.

Table 3. Calculated predictor functions using stepwise linear regression. For calculation, four objective measurements were used, which are abbreviated in the table with *Dur* (duration to build both target objects), *Pickup* (number of anticipatory pick up actions by experiment participant), *Utt* (number of utterances by experiment participant), and *Inst* (number of robot instructions).

Measure	Function	R^2	Significance	
Feelings	$324.68 + 0.77 * \mathcal{N}(\text{Dur}) + 27.35 * \mathcal{N}(\text{Pickup})$	0.27	Dur: $p \approx 0.16$	**Utt:** $p < 0.01$
	$-40.26 * \mathcal{N}(\text{Utt}) + 20.96 * \mathcal{N}(\text{Inst})$		Pickup: $p \approx 0.16$	Inst: $p \approx 0.13$
Intelligence	$405.02 + 0.58 * \mathcal{N}(\text{Dur}) - 18.70 * \mathcal{N}(\text{Utt})$	0.15	Dur: $p \approx 0.10$	**Utt:** $p < 0.05$
Behaviour	$487.33 - 10.96 * \mathcal{N}(\text{Pickup})$	0.12	Pickup: $p \approx 0.05$	
Task success	$447.74 + 0.40 * \mathcal{N}(\text{Dur}) - 17.54 * \mathcal{N}(\text{Utt})$	0.23	Dur: $p \approx 0.10$	**Utt:** $p < 0.01$

The calculated predictor functions show that all of the objective measurements influence user satisfaction in one way or the other:

- The number of user utterances has a strongly negative influence on the three categories *feelings of the user* (abbreviated with *Feelings* in table), *intelligence of the robot* (abbr. *Intelligence*), and *task success*. The duration to build both target objects had a slight positive effect in the same three categories.
- The number of anticipatory pick up actions by the user had a positive influence on category *feelings of the user* and a negative influence on category *robot behaviour*.
- The number of robot instructions had a strong positive influence on the category *feelings of the user*, but not on the other categories.

The R^2 values of this study are in the same range as the values of our previous user evaluations. However, the values are not as high as those reported in [11] and [6].

3.6 Discussion

The results of this study show an interesting correlation: we expected that the experiment participants will prefer the supportive robot over the instructive robot (see **H1**). However, the data suggests that the users accept both robot roles and simply take the counterpart in the interaction with the robot. This can be seen from the significant answers to the statements of the user questionnaire, where the users that worked with the supportive robot answered more positive to the statement "I knew what I could say or do at each point in the conversation". This indicates that the participants showed a more proactive behaviour themselves and followed the assembly plan more by themselves when the robot gave less instructions. This is in line with the work of Hinds et al. [4], who found that humans who work with a robot that takes the role of a supervisor, felt less responsible for the task.

In contrast to that, the users who worked with our instructive robot rated the statement "The robot gave too many instructions" lower than the users from the other group, which we interpret as confirmation for hypothesis **H2**: participants who worked with the instructive robot show a more passive behaviour. One of the objective measurements also supports this claim: users who worked with the supportive robot showed a proactive behaviour and executed anticipatory pick up actions significantly more often than users of the other group. These results are in line with research from cognitive psychology and cognitive neuroscience. For example [8] review a set of studies from these fields, which also prove that humans attune their actions when working together.

The results of the calculated predictor functions are not very surprising. However, it is interesting to note that the number of anticipatory pick up actions had a positive influence on the statements in the category *feelings of the user* and a negative effect on the category *robot behaviour*. The positive influence on the feelings of the user is a confirmation for hypothesis **H1**: the participants prefer to be proactive, thus a supportive robot fits better to their preferences. The negative effect of these measurements on the assessment of the robot's behaviour can be explained with robot errors during the interaction: when the robot made an error and for example gave the wrong instructions to the user or stopped working (which could happen sometimes during the experiments because of wrongly recognised construction pieces), the users had to pick up the pieces to finish building the target objects without getting instructions by the robot.

The number of user utterances also had a negative influence on the user satisfaction. This can be easily explained: in this experiment, the system was configured so that the users did not have to speak with the robot, as long as it performed well. The users only had to talk to the robot when they either did not understand the robot's utterances and had to ask for repetition or they needed to give a direct command to the robot to ask for a piece of the robot's workspace, which almost only happened when the robot made an error. Thus, the number of user utterances is a clear indicator for problems during the experiment.

4 Conclusion

The goal of this work was to research how humans react to different roles of a robot when they have to work with the robot on a common construction task. For that, we conducted an experiment in which a human and a robot together assemble target objects from a wooden toy construction set. In this experiment we programmed the robot to take different roles in the interaction. On the one hand, the robot took the role of an instructor and gave the humans instructions on how to build the target objects before helping them by handing over appropriate construction pieces. On the other hand, the robot took the role of a supporter that directly started handing over construction pieces to its human partner and only gave instructions when necessary. To our knowledge, there have been no similar experiments conducted yet to research the role of a robot in such a construction task.

We video-taped the experiment participants and analysed the automatically generated system log files to gather a set of objective measurements from the experiment. Additionally, we asked the participants to fill out a user questionnaire to get subjective measurements as well. The analysis of the gathered data showed that, in contrast to our expectations, the users did not prefer one of the two robot roles but simply took the counterpart to the role of the robot and adjusted their own behaviour to the behaviour of the robot. This was shown in one of the objective measurements as well as in the subjective ratings of the users: experiment participants picked up construction pieces significantly more often without the robot explicitly telling them when they worked with the supportive robot; additionally, users who worked with the instructive robot wanted to hear even more instructions although the robot already gave significantly more instructions to this experiment participant group. The analysis of the influence of the objective measurements on user satisfaction revealed that in the type of scenario that we presented here, users prefer to speak less, because spoken utterances were mainly used to resolve problems in the interaction.

In future work we want to research how humans perceive different roles of a robot in scenarios in which the robot interacts with more than one human. Furthermore, we plan to analyse the arm and head movements, gestures, and verbal utterances the robot can use to emphasize its own role in the interaction.

Acknowledgements. This research was supported by the European Commission through the projects JAST[1] (FP6-003747-IP) and JAMES[2] (FP7-270435-STREP). Thanks to Sören Jentzsch for help in annotating the video data.

References

1. Breazeal, C.: Social interactions in HRI: the robot view. Systems, Man, and Cybernetics, Part C: Applications and Reviews 34(2), 181–186 (2004)
2. Foster, M.E., Giuliani, M., Isard, A., Matheson, C., Oberlander, J., Knoll, A.: Evaluating description and reference strategies in a cooperative human-robot dialogue system. In: International Joint Conference on Artificial Intelligence (IJCAI 2009), Pasadena, California (July 2009)
3. Giuliani, M., Foster, M.E., Isard, A., Matheson, C., Oberlander, J., Knoll, A.: Situated reference in a hybrid human-robot interaction system. In: International Natural Language Generation Conference (INLG 2010), Dublin, Ireland (July 2010)
4. Hinds, P.J., Roberts, T.L., Jones, H.: Whose Job Is It Anyway? A Study of Human–Robot Interaction in a Collaborative Task. Human-Computer Interaction 19, 151–181 (2004)
5. Kipp, M.: Multimedia annotation, querying and analysis in ANVIL. In: Multimedia Information Extraction (2010)
6. Litman, D.J., Pan, S.: Designing and evaluating an adaptive spoken dialogue system. User Modeling and User-Adapted Interaction 12(2–3), 111–137 (2002)
7. Looije, R., Neerincx, M., Kruijff, G.J.M.: Affective Collaborative Robots for Safety & Crisis Management in the Field. In: Intelligent Human Computer Systems for Crisis Response and Management (ISCRAM 2007), Delft, Netherlands (May 2007)

[1] http://www6.in.tum.de/Main/ResearchJast
[2] http://www.james-project.eu

8. Sebanz, N., Bekkering, H., Knoblich, G.: Joint action: bodies and minds moving together. Trends in Cognitive Sciences 10(2), 70–76 (2006)
9. Tapus, A., Mataric, M.: Socially assistive robots: The link between personality, empathy, physiological signals, and task performance. In: AAAI Spring, vol. 8 (2008)
10. van Breemen, A.J.N.: iCat: Experimenting with animabotics. In: AISB 2005 Creative Robotics Symposium (2005)
11. Walker, M., Kamm, C., Litman, D.: Towards developing general models of usability with PARADISE. Natural Language Engineering 6(3–4), 363–377 (2000)

People's Perception of Domestic Service Robots: Same Household, Same Opinion?

Julia Fink, Valérie Bauwens, Omar Mubin, Frédéric Kaplan, and Pierre Dillenbourg

Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland
{julia.fink,valerie.bauwens,omar.mubin,frederic.kaplan, pierre.dillenbourg}@epfl.ch

Abstract. The study presented in this paper examined people's perception of domestic service robots by means of an ethnographic study. We investigated initial reactions of nine households who lived with a Roomba vacuum cleaner robot over a two week period. To explore people's attitude and how it changed over time, we used a recurring questionnaire that was filled at three different times, integrated in 18 semi-structured qualitative interviews. Our findings suggest that being part of a specific household has an impact how each individual household member perceives the robot. We interpret that, even though individual experiences with the robot might differ from one other, a household shares a specific opinion about the robot. Moreover our findings also indicate that how people perceived Roomba did not change drastically over the two week period.

Keywords: attitudes towards robots, domestic service robots, human-robot interaction.

1 Introduction

Within the last years several service robots for personal and domestic use, such as vacuum cleaning robots, lawn mowing robots and toy robots have been introduced into the mass market. The number of domestic service robots deployed in homes increases constantly but lags behind early estimations. According to Bill Gates' article "A robot in every home", it seems that for the personal service robot industry, a similar scenario is going on as we had for the computer business 30 years ago [1]. On the one hand, there is a lack of common standards and platforms, so that robot developers usually have to start from scratch when building new robotic devices. On the other hand, when developing robots for domestic use, we cannot deny social implications of human-robot interaction, human behavior and people's expectations. It has been shown that social aspects play a crucial role in technology adoption [2] and that people tend to perceive artifacts showing intentional behavior as characters or even creatures [3]. The tendency to anthropomorphize nonhuman agents also holds for technical tools or robots that seem to lack capabilities of having social interaction [4]. An ethnographic study conducted by Forlizzi et al. revealed that a vacuum cleaning robot in contrast to a traditional vacuum cleaner deployed in the home changed people's cleaning activities and how they used other tools [5]. However, still little is known about how people actually use domestic service robots and how they react to

B. Mutlu et al. (Eds.): ICSR 2011, LNAI 7072, pp. 204–213, 2011.

them. The majority of literature on human-robot interaction in this area is of technical nature. But if we aim to develop and introduce into homes human-oriented robots that perform tasks for people in their everyday environment and safely co-exist with them, we need to understand not only technological affordances but also social aspects of human-robot interaction. In addition to more general issues about how people perceive domestic robots, questions arise about how people, children, the elderly and pets react to a robot in the home, what different kinds of difficulties they face, how they explore the robot and whether and how they integrate it into their daily life.

To address these questions, we gave a vacuum cleaning robot to nine households and followed people's experiences with the robot over two weeks. We decided to use iRobot's™ Roomba vacuum cleaning robot to study people's perception of it in homes because it has been available as a consumer product for several years. The robot is fully developed, quite robust and financially affordable. Surprisingly enough, even though, according to iRobot™, about 6 million units have been sold already, yet little is known about how people accept it. However, we argue that, acceptance of a domestic robot is not only related to the individual perspective but also to a certain 'household perspective' (e.g. a 'family perspective', such as shared values and beliefs). Thus we claim that, being member of a specific household (e.g. a family or a couple living together) could influence the acceptance and how people perceive the service robot in their home. To verify this, we analyzed people's attitudes and their reactions to the robot not only on an individual but also on a household level. A holistic view that regards people not only as users but rather as social actors within the ecology of their home will help to understand human-robot interaction in domestic settings [5, 6, 7]. We explored people's perceptions and examined how their reactions evolved. We visited each of the nine household twice, which made in total 18 semi-structured interviews. We collected quantitative data in form of a questionnaire with rating scales filled out by all household members (n=26) as well as qualitative data through interview conversations, home tours and on-site observations.

2 Related Work

In this part, we situate our project with respect to related work that studied the acceptance of technology and human-robot interaction in domestic spaces. Some studies have been carried out that focused on user needs for (future) domestic robots [8, 9]. Participants of Scopelliti's study did not have a clear idea about what a future domestic robot could be or do in the household but rather responses seemed to emerge from science fiction movies or novels about robots [8]. The authors report some significant gender and age differences in the perception of domestic robots, for instance, in terms of confidence in the capabilities of robots, emotional reactions to a domestic robot, preferred characteristics and interaction modalities of robots. Based on the assessment of user needs, Sung et al. gave several suggestions in terms of design, namely that a domestic robot needs to provide a certain amount of human control, be compatible with the user's domestic environment, and take gender into consideration. In terms of social relation to the device, participants preferred a friendly designed robot that would act as a professional butler but not as a friend. Dautenhahn reported a similar finding and stated that people want to view their home robots as machines, assistants and servants, performing various tasks [10].

Several researchers studied social relationships to home technologies and domestic robots, such as intimacy, affective quality, and emotional attachment as well as their role during the process of adoption [11, 2]. Social effects of technology adoption have been described as 'intimacy in computing' [3] or 'intimate ubiquitous computing' [12, 2] and despite using different technologies draw a similar conclusion: intimacy leads to greater acceptance of technology and perceived usability. It is likely that these effects also apply to robots such as Roomba [6, 10, 13]. It has been observed that domestic robots did not only lead people to change their cleaning patterns or the physical arrangement of their home but that they also developed relationships with the robot by assigning a name or ascribing personality traits to the device [6]. Intimacy can inform device adoption and help people to manage for example unreliability [13]. By feeling socially attached to the robotic vacuum cleaner, people were able to derive increased pleasure from cleaning, and expended effort to fit the robot into their homes. Scopelliti et al. concluded that acceptability of robotic devices in home settings, especially for elderly people, would not only depend on the practical benefits of the device, but on complex relationships between the cognitive, affective and emotional components of people's images of robots [8].

One concept often referred to when talking about the acceptance of robots is anthropomorphism. Epley et al. describe anthropomorphism as the tendency to "imbue the real or imagined behavior of nonhuman agents with humanlike characteristics, motivations, intentions or emotions" [4]. One of the derived determinants the authors name to explain when people are likely to anthropomorphize is people's desire for social contact and affiliation (sociality motivation). Anthropomorphism can thus play an important role concerning the acceptance of robots. It may serve as an effective method for improving usefulness and it has been shown that people were more likely to cooperate and work with a playful robot than with a serious robot [14].

Besides social relations between humans and robots, researchers also studied interactions between technological and social space in the home during the process of technology adoption: Venkatesh stressed the role of social space (e.g. family members, household activities, personal needs etc.) influencing the technological one [15]. He claimed that a thorough understanding of technology adoption in the household requires a theory of household behavior. These issues have more recently been studied by [16] for whom the composition of a household influences how people use a domestic service robot. The results of her survey study are based on the responses of more than 350 Roomba users.

In summary, the reviewed literature suggests that, for investigating service robots in domestic settings, it is worth considering not only users as individuals and looking for gender or age differences but also take into account the characteristics of the entire household. This hypothesis builds on the fact of a shared "physical" and "social space" in a domestic environment as well as on findings from previous studies showing that the household composition affected the expectation of Roomba as a practical tool [7]. Further, we assume that the adoption of a domestic robot in the context of the home not only happens as an individual but as a whole household [11]. In our study, we therefore apply a holistic household view and quantitatively argue for doing so.

3 Study Design

3.1 Method and Measurements

Our study aimed at exploring people's perception of a domestic robot they have in their home for two weeks. We combined quantitative and qualitative research methods which enabled us to deeply explore participant's beliefs and reactions to robots. A part of the methodology of our approach was similar to that followed by Sung et al. and we adapted a questionnaire from her research [17].

Following a user-centered approach, we implemented a recurring questionnaire into several semi-structured interviews with the participants. The interviews took place at each participant's home and lasted about 1.5 hours each. They were audio recorded and qualitatively re-transcribed. We asked all household members to be present during our household visits; however, this was not always the case. To enhance discussion and examine social roles in the household, participants of the same household were interviewed collectively, thus children were not separated from their parents, for instance. In total, each of the nine households was visited twice during the two weeks which makes in total 18 visits. In addition to that, we have conducted a preliminary interview one week prior to the study, including a home tour. We got to know the household, the people and pets living in the household, asked them about their attitude towards technology and robots as well as their cleaning routine and the structure of their daily life. For the following two visits we took the questionnaire with us each time and asked participants to rank Roomba on a seven point Likert scale (1-7) in respect to its intelligence, usefulness, ease of use, experienced fun, expected impact on the household and the participant's overall impression. A questionnaire with Likert rating scales to assess people's attitude towards robots has also been used by [8, 17].

Participants filled out the questionnaire three times:

- Before they had seen Roomba (T1), in order to investigate their expectations and imaginations of the vacuuming cleaning robot;
- After they had unpacked and interacted with Roomba for about 10 to 15 minutes (T2) to catch their initial impression;
- Two weeks after they received Roomba (T3).

The participants' ratings on the questionnaires were filled in a spreadsheet, regarding one decimal steps where two coders agreed on the exact interpretation of each mark (e.g. if the mark was understood as 5.4 or 5.5).

3.2 Participants

Our sample consisted of nine households with a total of 30 participants. Four of the fifteen children were six years and younger, so we did not ask them to evaluate the robot using the questionnaire. 26 participants filled out the questionnaire: 14 men and 12 women, thereof 11 children, and 15 adults, ranging from 7 to 70 years old. The mean age of all participants was 30, of all adults older than 18 years the mean age was 43. More specifically, we had three single-headed (one single parent) and six double-headed households. Six households had children from six months to 18 years old.

Finally, since our study focused on cleaning, we recruited households with pets (n=5), such as dogs and cats, and households with cleaning services (n=4). Among the four who hired cleaning services, three received the service once a week and one every other week. Everyone of the quite heterogeneous sample lived in the area of Lausanne in the French speaking part of Switzerland. With more than one third of its inhabitants being foreigners, Lausanne displays multi-faceted cultures besides a Swiss-French mentality. Our sample reflects this cultural diversity including people of American, Austrian, British, Danish, French, German, and Swedish background besides two Swiss households. Interviews and questionnaires were carried out in English, French or German, depending on the language all household members spoke fluently.

3.3 Design

Our previously mentioned questionnaire has been used to evaluate the perception of Roomba by the households. The quantitative analysis was divided in three parts. The first analysis was run as a simple repeated measures ANOVA with time as the independent variable. The second analysis included a number of between subject factors that were also incorporated in a mixed design ANOVA. For the first two parts of the analysis we considered every individual household member as an individual data point. In the third part of our analysis we aggregated all the members of one household to make a single data point and a repeated measure ANOVA was done. In the third part the single data point would represent the household opinion (which for the case of our two single households displayed the opinion of a single person).

4 Findings

Our overall aim was to find out how people and households perceive a domestic robot in their home. This section is structured in two parts. First, we present the overall results of how our participants on an individual level perceived the robot in respect to its intelligence, usefulness, usability, fun, attachment, impact, and their overall impression. In the second part we present our findings on a household level aggregating data of individual members to verify our assumption that response behavior is influenced by the fact of being part of a specific household. Although in general how people rated Roomba in terms of the provided topics (Table 1) didn't change significantly over time, how they qualitatively described the robot changed with time. These qualitative results will be reported elsewhere.

4.1 Individual Level

We first present the quantitative analysis on an individual level over time. Table 1 shows the means and standard deviations of each topic over the three time points. Note that for some individuals we did not have ratings for all three occasions. Therefore they had to be dropped from the ANOVA analysis. Thus, our sample size was 19.

Table 1. Means and standard deviations for each topic over the three times. Repeated measures ANOVA with time as within subjects factor.

Topic	T_1 Mean	T_1 Std dev	T_2 Mean	T_2 Std dev	T_3 Mean	T_3 Std dev	F (2, 36)	p
Intelligence	4.38	1.68	4.63	1.7	4.22	1.71	0.70	0.50
Usefulness	5.2	1.39	5.44	1.09	5.0	1.36	0.81	0.45
Ease of use	5.74	1.09	6.28	0.63	6.02	1.25	1.59	0.22
Fun	4.38	1.70	4.43	1.7	4.43	1.72	0.01	1.0
Attachment	3.8	1.43	4.28	1.88	4.23	1.71	0.92	0.41
Impact	4.26	1.74	4.56	1.59	4.17	1.29	0.62	0.55
Overall impression	5.43	1.21	5.56	1.21	4.95	1.29	1.28	0.29

We executed a repeated measures ANOVA with time as the within subjects factor. As evident by Table 1, time did not have a significant effect on any of the topics. There is however an interesting pattern that for all of the seven topics, the T2 mean value is the maximum value.

After testing the main effects we aimed to determine the impact of any external factors or biases on an individual's rating. The factors were considered one at a time because if we had included them together in a mixed design ANOVA this would have reduced the sample size even further for every combination.

We first wished to examine any gender effects. Surprisingly, gender did not have a significant effect on the ratings (see Table 2). We would have expected male participants giving more positive ratings than female participants purely based on the fact that Roomba is a technological device. However, in our qualitative observations it turned out that female householders were the primary users of Roomba and this fact could have resulted in females giving equally high ratings as compared to men. This goes along with results of [10, 16]. However, other studies reported significant gender differences [8]. We further analyzed variances for different age groups and pet owners but could not find any significant results. This might be due to the rather small size of the sample within each age group.

Table 2. Mixed design ANOVA with time as within subjects factor and gender as between subjects factor. (f=female; m=male)

Topic	T_1 m Mean	T_1 f Mean	T_2 m Mean	T_2 f Mean	T_3 m Mean	T_3 f Mean	F (1, 17)	p
Intelligence	5.07	3.62	5.02	4.19	4.68	3.72	2.82	0.11
Usefulness	5.81	4.52	5.84	4.99	4.95	5.00	3.13	0.10
Ease of use	6.26	5.17	6.39	6.17	5.93	6.11	1.52	0.24
Fun	5.05	3.63	4.53	4.31	4.64	4.19	1.24	0.28
Attachment	3.49	3.92	4.14	4.21	3.81	4.69	0.58	0.46
Impact	3.67	4.26	4.59	4.39	4.34	3.99	0.00	0.98
Overall impression	5.75	4.86	6.00	5.10	4.82	5.10	2.89	0.11

4.2 Household Level

The next factor was family code which indicated to which particular household an individual belongs. We executed a mixed design ANOVA with time as the within subject factor and family code as the between subject factor. The results are summarized in Table 3.

Table 3. ANOVA results showing the effect of family code on the ratings

Topic	F (1, 10)	p
Intelligence	2.68	0.07
Usefulness	3.9	0.02*
Ease of use	1.89	0.16
Fun	9.77	0.02*
Attachment	9.83	0.02*
Impact	6.45	0.16
Overall	1.88	0.27

The results of the ANOVA showed that for some of the topics, being part of a particular household had a significant effect. As an example for this, Figure 1 shows individual responses for each member of household F9. Consisting of four household members (single father and three boys from eight to eleven years old) the graph illustrates how the individual ratings form a certain family opinion, where in this case, one family member (P19) differed in his responses from the others. For this family, there was clearly no peak at T2.

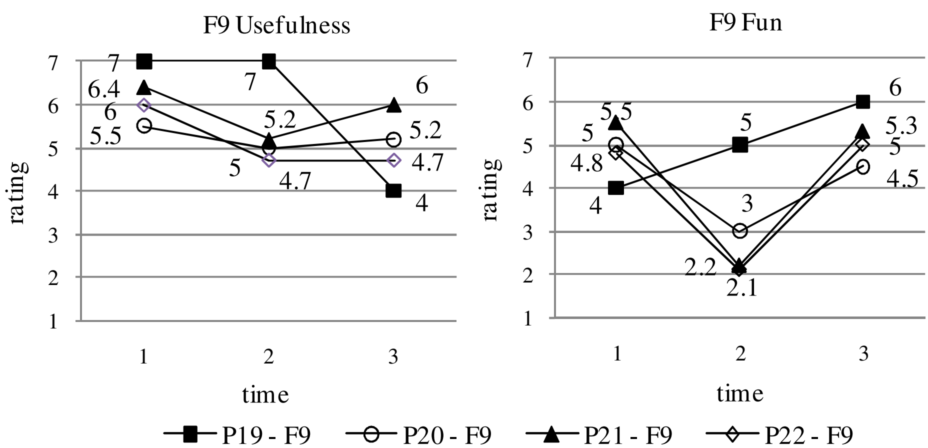


Fig. 1. Individual responses for usefulness and fun for each of the four members of a family

We wanted to examine in more detail how belonging to a family influences people's response behavior and carried out further analyses by aggregating the scores of the individual members of a single family. We carried out a repeated measures ANOVA on the aggregated scores of the family with time as the within subject (family) factor. However, we did not get any significant results for this case.

In conclusion, our data shows that during two weeks, on an individual level, there have not been any significant changes in how people rated their domestic robots in respect to intelligence, usefulness, ease of use, fun, attachment, impact and the overall impression. However, the rating of various topics was significantly influenced by the fact that participants belonged to the same household. Thus, on a household level, we can report significant effects on people's perception of robots. Another interesting pattern that emerged was that for individuals we could see a peak effect. The mean rating for time T2 was higher than T1, but T3 was always less than or equal to T2.

5 Limitations and Discussion

Our aim was to study human-robot interaction including social aspects in a setting that was as natural as possible. Conducting qualitative research in a domestic site raises various challenges, such as to capture reliable data in a highly uncontrolled environment with constraints of privacy and temporality in the home. Furthermore, in our study responses might be biased due to the fact that people received a cleaning robot from us and that they did not have to make a financial investment with buying the device. This might have influenced how people perceived the robot.

The interpretability of our results is limited by the sample size counting nine households with 30 participants but several missing responses. However, a study on a larger scale would not have been realizable due to time and resource constraints. We nevertheless analyzed data on a household level and hypothesized that being part of a particular household influences people's perception of robots. It remains to mention that two of the nine households consisted of one person only. In addition to the heterogeneity of participants might have contributed to the observed effect. Cultural differences in the perception of robots have been investigated by [22].

As mentioned before, this study is part of a longitudinal ethnography. It documents the evolution of people's perception of robots for only two weeks. The responses are likely to portray certain novelty effects which can be described as the first responses to a technology [17, 19]. Covering a period of two weeks, data does not allow drawing conclusions in terms of usage patterns or continuing adoption of the device. Long-term usage of robots and usage patterns beyond novelty effects have been examined by [7, 17, 19, 20]. We didn't find a significant change in how people overall rated the robot during the two weeks and assume that amongst others, novelty factors might be one reason for this. Literature suggests that people's expectations are quite powerful in shaping the initial experience [6] and this in turn seems to be crucial for the further process of adoption [11]. Another reason why the perception did not change drastically over time might be that Roomba could have been perceived more as a technical device rather than as a social robot. Roomba lacks the ability to learn and monotonously performs the task of vacuuming. It remains to see how far people's rating of Roomba changes over a longer period of time or whether with Roomba people's perception remains quite stable after people have formed first opinions.

We presented findings about how people perceive a domestic robot on an individual level as well as on a household level and hypothesized different effects.

- On an individual level we did not find significant gender or age effects. This was surprising but might be explained through results on the household level.
- On a household level we found a significant effect that we called 'family effect' and speculate about their characteristics in terms of the shared social values (H1), the shared physical environment (H2) or an opinion leader effect (H3).

The observed 'family effect' can be impacted by the fact that the members of one household have not been separated during the interviews. More important, the effect may result from social norms, values and beliefs that are shared in one household (H1) but also be due to the physical layout of the home (H2). The fact that an apartment has stairs or an open kitchen may have an effect on the relevance of Roomba, for instance. Similarly, Severinson-Eklundh argues that understanding human-robot relationship includes consideration of the group of people involved, their social norms, roles and beliefs, as well as the shared physical environment [18]. Besides this, we saw that, while families rated Roomba, there tended to be an opinion leader (H3) who influenced the others. In social science, group conformity and decision making has been extensively studied since the 1950ies (e.g. Asch's experiment on group conformity). It will be important in further studies to try to disentangle the various effects we encounter in domestic environments.

Acknowledgements. We thank all our participants for their time and engagement in this study. We also thank iRobot™ and iRobotics GmbH for their support. Further, we thank JaYoung Sung and researchers from Georgia Tech for sharing their experience with us. This research was supported by the Swiss National Science Foundation through the National Centre of Competence in Research on Robotics.

References

1. Gates, B.: A Robot in Every Home. In: Scientific American, pp. 58–65 (January 2007)
2. Venkatesh, V.: Determinants of Perceived Ease of Use: Integrating Control, Intrinsic Motivation, and Emotion into the Technology Acceptance Model. Information Systems Research 11(4), 342–365 (2000)
3. Reeves, B., Nass, C.: The Media Equation: How People Treat Computers, Television, and New Media like Real People and Places. Cambridge University Press, New York (1996)
4. Epley, N., Waytz, A., Cacioppo, J.T.: On seeing human: A three-factor theory of anthropomorphism. Psychological Review 114(4), 864–886 (2007)
5. Forlizzi, J.: How Robotic Products Become Social Products: An Ethnographic Study of Cleaning in the Home. In: HRI 2007 Proceedings, Arlington, VA, USA, pp. 129–136. ACM Press (2007)
6. Forlizzi, J., DiSalvo, C.: Service Robots in the Domestic Environment: A Study of the Roomba Vacuum in the Home. In: HRI 2006 Proceedings, Salt Lake City, UT, USA, pp. 258–265. ACM Press (2006)
7. Sung, J., Grinter, R.E., Christensen, H.I.: Domestic Robot Ecology. International Journal of Social Robotics 2(4), 417–429 (2010)

8. Scopelliti, M., Giuliani, M.V., Fornara, F.: Robots in a domestic setting: a psychological approach. Journal Universal Access in the Information Society 4(2), 146–155 (2005)
9. Sung, J., Christensen, H.I., Grinter, R.E.: Sketching the Future: Assessing User Needs for Domestic Robots. In: IEEE International Symposium on Robot and Human Interactive Communication, Toyama, Japan, pp. 153–158 (2009)
10. Dautenhahn, K., Woods, S., Kaouir, C., Walters, M.L., Koay, K.L., Werry, I.: What is a Robot Companion – Friend, Assistant or Butler? In: IEEE Robotics Society of Japan International Conference on Intelligent Robots and Systems (2005)
11. Rogers, E.M.: Diffusion of Innovations. Fifth edn. Free Press, New York (2003)
12. Bell, G., Blythe, M., Gaver, B., Sengers, P., Wright, P.: Designing Culturally Situated Technologies for the Home. In: CHI 2003 Proceedings, Ft. Lauderdale, FL, USA, pp. 1062–1063. ACM Press (2003)
13. Sung, J., Guo, L., Grinter, R.E., Christensen, H.I.: "My Roomba Is Rambo": Intimate Home Appliances. In: Krumm, J., Abowd, G.D., Seneviratne, A., Strang, T. (eds.) UbiComp 2007. LNCS, vol. 4717, pp. 145–162. Springer, Heidelberg (2007)
14. Kiesler, S., Goetz, J.: Mental models of robotic assistants. In: CHI 2002 Proceedings, Minneapolis, MN, USA, pp. 576–577. ACM Press (2002)
15. Venkatesh, A.: Computers and Other Interactive Technologies for the Home. Communications of the ACM 39(12), 47–54 (1996)
16. Sung, J., Grinter, R.E., Christensen, H.I., Guo, L.: Housewives or Technophiles?: Understanding Domestic Robot Owners. In: HRI 2008 Proceedings, Amsterdam, The Netherlands, pp. 129–136. ACM Press (2008)
17. Sung, J., Christensen, H.I., Grinter, R.E.: Robots in the Wild: Understanding Long-term Use. In: HRI 2009 Proceedings, La Jolla, CA, USA, pp. 45–52. ACM Press (2009)
18. Severinson-Eklundh, K., Green, A., Hüttenrauch, H.: Social and collaborative aspects of interaction with a service robot. Robotics and Autonomous Systems 42(3-4), 223–234 (2003)
19. Hüttenrauch, H., Severinson-Eklundh, K.: Fetch-and-carry with CERO: Observations from a long-term study with a service robot. In: IEEE International Workshop on Robot and Human Interactive Communication, Berlin, Germany, pp. 158–163 (2002)
20. Koay, K.L., Syrdal, D.S., Walters, M.L., Dautenhahn, K.: Living with Robots: Investigating the Habituation Effect in Participants' Preferences During a Longitudinal Human-Robot Interaction Study. In: IEEE International Conference on Robot and Human Interactive Communication, Jeju, Korea, pp. 564–569 (2007)
21. Venkatesh, A., Kruse, E., Chuan-Fong Shih, E.: The networked home: an analysis of current developments and future trends. Cognition, Technology, and Work 5, 23–32 (2003)
22. Kaplan, F.: Who is afraid of the humanoid? Investigating cultural differences in the acceptance of robots. International Journal of Humanoid Robotics 1(3), 1–16 (2004)

RoboCup@Home: Adaptive Benchmarking of Robot Bodies and Minds

Tijn van der Zant[1] and Luca Iocchi[2]

[1] Cogitive Robotics Laboratory, A.I. dept, faculty of Natural Sciences, University of Groningen
tijn@ieee.com

[2] Department of Computer and Systems Science, University of Rome "La Sapienza"

Abstract. RoboCup@Home is the largest benchmarking effort for domestic service robots. The benchmarking is in the form of a competition, with several yearly local competitions and an international one. Every year the tests become more complex, depending on the results of the previous years. In the past four years the focus has been on benchmarking physical aspects of the robots, such as manipulation, recognizing people and human-robot interaction. In 2010, for the first time, there is a test which is targeted at the mental cognitive capabilities of the robot. In order to guarantee scientific quality of the proposed solutions and effective integration in a fully working system, all the tests include different capabilities and change every year. This novel feature of RoboCup@Home benchmarking raises the question of: How can effective benchmark tests be defined and at the same time measure the progress over many years? In this paper we present the methodology applied in and results of RoboCup@Home for measuring the effectiveness of benchmarking service robots through competitions and present a new integrated test for benchmarking the cognitive abilities of a robot.

1 Introduction

RoboCup@Home[1] is the largest human-robot interaction and domestic service robot benchmarking effort. The goal of RoboCup@Home is to foster the development of a versatile domestic service robot that can operate safely in all situations that people encounter in daily life. RoboCup@Home [5,6] is part of the RoboCup Federation [3,4], which has been promoting, for many years already, the use of competitions in order to drive research towards robust techniques and useful applications and to stimulate teams to compare their approaches on a set of common testbeds. This dynamical form of benchmarking has dramatically improved the standards of robotics. Robots have become much more robust and reactive.

RoboCup@Home competition focuses on the benchmarking of domestic service robots. It consists of several tests where many functionalities are tested at the same time in an integrated way. Most of the tests integrate navigation, localization, object recognition, manipulation, speech recognition, etc. Examples

[1] http://www.robocupathome.org/

B. Mutlu et al. (Eds.): ICSR 2011, LNAI 7072, pp. 214–225, 2011.

of tests are: the robot has to find and bring objects scattered in the environment in unknown locations, assist with cooking or bringing a drink, follow its owner through a real world environment such as a shopping mall, recognize known persons in a confusing situation, reason about the world and assist with shopping in a real supermarket. Most tests are performed in an apartment scenario and some tests are performed in the real world. RoboCup@Home started in 2006 and has now many local events and a global one. This year (2011) 24 teams have participated in the global event. During this event more than 30 robots are tested resulting in over 175 separate benchmarks. It is the most variable benchmarking effort within the RoboCup Federation, and arguably the most versatile robotic benchmarking effort in general. This versatility is both the strength and the weakness of the competition.

Benchmarking has many advantages, such as the creation of standard references and metrics. However, there is also a possible disadvantage when benchmarking remains static over time, that is the progress of solutions towards a local optimum, without having a guarantee that the devised solution will work in general. In robotics this can be a huge problem. By changing the tests on a yearly basis, it is possible to ensure that specialized solutions looking for a local optimum will not be rewarded in the long term.

On the other hand, a dynamic benchmark that changes over time introduces the problem of relating the different generations of the particular benchmark [2]. This problem is acute regarding tests that require the robot to operate in the real world. Statistical analysis on the results is the scientific approach for this. Robots have to solve not a single task, but many tasks, where capabilities are tested several times in different settings.

The downside of the changing tests is that this co-evolutionary process does not guarantee progress into the desired direction of versatile autonomous domestic service robots. It is essential that scientific standards are applied to this type of flexible benchmarking. In RoboCup@Home statistical analysis of the outcome of the competition is used to steer the league into interesting directions.

This paper thus presents an important contribution to benchmarking human-robot interaction, domestic service robots and robot cognition. In the following the methodology used to drive the benchmarking through competitions is presented, and a new test aiming at measuring cognitive abilities of a robot is described. This is, to the best of our knowledge, the first general benchmark for high-level reasoning of cognitive robots. Also the results of the analysis performed during the past 4 years are presented, showing the effectiveness of the benchmarking activity.

A more detailed description of the RoboCup@Home competition, the tests used until 2008 and past results can be found in [8]. In this paper, in addition to updating the results with 2009 and (partially) 2010 competitions, the methodology for steering the dynamic benchmarking is desribed in more detail. Also, a novel test for the evaluation of the high-level cognitive abilities abilities of robots is introduced.

2 Defining the Road Map Through Adaptive Benchmarking

The road map of RoboCup@Home is not completely clear. The goal is clear (a versatile domestic service robot) but how to get there is not. This is due to the fact that the prediction of future development in robotics hardware and software is not easy. This uncertainty does not prevent benchmarking though, but the approach has to be adapted to the response of the research groups participating to it. In this situation, a fixed setting in benchmarking is not likely to bring good solutions, since it may be either too complex or too simple. Moreover, defining several benchmarks is very complicated and does not ensure a large participation of research groups. The ideal solution is to devise flexible and variable benchmarks that reflect the lessons that are continuously being learned.

An example we report from RoboCup@Home experience is the case of speech and gesture recognition, where one main goal is to understand an unknown (English speaking) person. Since the start of the @Home competition, it has been mandatory to only use speech or gestures to interact with the robot. At the international competition though, there is a tribune with up to several hundreds of spectators making a rather large amount of noise. In the first year the standard solution was to use wireless headsets to interact with the robot. The next year bonus points were introduced for not using a headset. Soon several teams tried to gain points using on-board microphones [1].

Since effective and robust solutions for speech recognition have been developed by almost all teams so far, gesture recognition was not considered. In 2009 we introduced a test where gesture recognition was stimulated by means of being able to get points for it explicitly. Gestures were restricted to hand and arm movements to avoid solutions where people have to perform unnatural actions. One team got the full score for this, demonstrating that it is possible. HRI will be tested in a real noisy environment: a shopping mall including actual customers. During the new 'follow me' test the user has to walk 3 meters away while the robot is standing still, and then give a command to make the robot move towards the user. The noise levels in such an environment will make this task very challenging and it is unclear to us whether speech or gesture will be more reliable, probably the right combination of both will be the best choice. With many teams attempting to find an effective and robust solution in the very challenging environments of RoboCup@Home, we will obtain some important results:

1. Statistical evidence of performance of the developed techniques is collected
2. The difficulty of the task is measured
3. Based on the results, recommendations for changing the tests in 2011 are discussed amongst peers

We firmly believe that this approach to benchmarking will quickly raise the general level of research groups developing solutions in domestic service robotics and may even fill the gap between laboratory experimental settings and real robotic applications, thus providing an important link with industries. The statistical benchmarking approach is the only viable approach for the benchmarking of real-world robots. Any form of fixed benchmarking will only create local optimal solutions that are unlikely to work in the real world. This approach could be effectively ported to other robotic application fields where the environment has a high level of uncertainty, such as the fields of rescue, space and reconnaissance robots.

3 From Physical to Mental Capabilities

Benchmarking physical capabilities, such as Self Localization And Mapping, manipulation of objects and recognizing persons, is not an easy task. The capabilities have to be tested in different environments and situations. RoboCup@Home provides these environments and situations. In the first years the competition was situated in an apartment, which is unknown beforehand and looks different at every competition. During the competition random changes are made to the apartment to simulate a place where people live and which they adapt to their likening. The setting is different for every test. It could be that a person has lost an object and the robot has to find it, that the robot has to remember and serve the correct drinks to guests or assist with shopping in a supermarket.

In 2010 new environments have been introduced. There is a test where the robot has to follow its user through a setting in a public area. During this test the user will be occluded from sight, walk away several meters and stand close to other persons in order to test the quality of the following behaviors of the robot. Another environment is used in the supermarket test. In this test the user and the robot go into a real supermarket. The robot has to assist the user by getting objects (such as a box of cornflakes) from a shelf and hand it over to the user. Also, at the check-out, the robot has to go back into the supermarket and get an item that the user 'forgot'. This test could lead to multi-user robotic applications that can assist customers with physical challenges. In the future the environments will be extended to, for example, going into town using the public transport to do some shopping or to accompany children walking from school to home.

What's missing in the test description mentioned so far is the testing of mental or high-level cognitive capabilities. Although the robots need to posses some form of intelligence for these tasks, it is important that the robots are ready for new and unforeseen situations. During the competitions there is not enough time to benchmark all possible situations which might occur in real life. The only solution is to test the cognitive capabilities of the robots, in order to assess whether it is likely that the robot can handle new situations, and use statistics for extrapolation of the results.

3.1 The First Cognitive Benchmark in Robocup@Home

In this section we present an integrated test to measure the cognitive capabilities of a robot. This test is called 'General Purpose Test' and it focuses on the following aspects.

– There is no predefined order of actions to carry out. This is to slowly get away from state machine-like behavior programming.
– There is an increased complexity in speech recognition compared to the other tests. Possible commands are less restricted in both actions and arguments. Commands can include multiple objects, e.g., "put the mug next to the cup on the kitchen table"
– The test is about how much the robot understands about the environment and aims for high-level reasoning.

The task will be executed as follows. The robot enters the arena by driving to a specified location. The referees select an action from the list of possible actions and command the robot to perform the desired task. Once the robot has successfully solved or interrupted the execution of the task, a new action is selected. The robot can carry out a maximum of three actions within 10 minutes. Actions, locations and objects are taken from previous tests. Since this is an advanced test in the competition, we can correctly assume that participating teams are familiar with such actions, locations and objects and have already solved in an effective way the basic functionalities, like navigation, SLAM, object recognition, manipulation, HRI, etc. This guarantees that the main focus of the test is on high-level cognitive capabilities.

3.2 Actions and Categories of Tasks

Regarding the following specifications of actions and complexity classes, several actions can be combined for a *compound action*, for example,

– "This is John, now follow John"
– "Go to the table and point at the yellow cup"
– "Go to the living room and find John"
– "This is a coffee mug (showing the mug), now go to the kitchen and point at the mug on the kitchen table"

Moreover, we have defined three different categories of tasks that form the test.

1. **Understanding orders:** Understand complex sentences and perform the correct actions:
   – Robot, go to the bed room, pick up the red cup next to the bed and put it on the kitchen table
   – Robot, follow the person in front of you and stop when you are in the kitchen to pick up the red cup from the table
   – Robot, go to the living room, find John (a known person), tell him your name and then come back to me

2. **Understanding itself:** The robot gets a simple command, but it does not have all the information necessary to complete the task. The robot may ask up to five questions to get relevant information, before starting with the task. Examples are the following:
   – *user: "Robot, bring me a cup"*
     robot: "Which cup?"
     user: "My cup." or "The red cup."
     robot: "Where is the cup?"
     user: "On the table."
   – Another example:
     user: "robot, put the red cup on the table!"
     The robot does not have the red cup in its hands, it needs to get it. Since searching takes a long time, it can ask:
     robot: "Excuse me, do you know where the cup is?"
     user: "It's in the bedroom."
     And the robot is allowed to pursue further:
     robot: "Could you tell me where exactly in the bed room?"
     user: "It's on the table next to the bed"
     And the robot performs the action
3. **Understanding the world:** The robot has to answer 3 questions about (events in) the world. The necessary recognition capabilities are all from previous tests. The first two questions are about understanding the world as it is, for example asking how many persons there are in the room. The third question is where the robot is being *"tricked"*. Examples of a trick are putting a box over a cup and asking the robot where the cup is, or having a person sitting at the table with the robot and a person standing and asking the robot how many people are sitting at the table. More specific examples are the following:

   (a) The robot is in front of a table with several cups in different colors on it. The human can ask and do the following:
   user: "Robot, can you tell me what's on the table?", and the robot has to answer something like
   robot: "I see three cups, one is red, one is blue and one is yellow" or
   robot: "I see a red cup, a blue cup and a yellow cup".
   A wrong answer would be that the robot says that it only sees one cup, for example.

   (b) Then the human can ask:
   user: "Robot, where is the red cup in relation to the other cups?", the correct answer would be
   robot: "The red cup is left compared to the other cups.", of course depending on the position of the cup.

   (c) A third action/question could be that the human takes away 2 of the cups, and the puts a box over the remaining (blue) one. The human can then asks:

user: "Robot, do you know where the blue cup is?", the correct answer would be
robot: "The blue cup is under (behind) a box",
a wrong answer would be "There is no blue cup". The difference is between stating what the robot sees, or testing that the robot is aware of object persistence. The robot is in front of a table with several cups in different colors on it. The human can ask and do the following:

Although this task is complex, the criterion is always the same: Did the robot show understanding, or is it only stating what it sees.

The setup is experimental since this is the first year that high-level cognition is tested. The test is loosely based on skills which are still difficult for robots in general (such as parsing a complex sentence into a set of actions or reasoning about its own capabilities) and developmental psychology, such as reasoning about counting objects and occlusion events [7].

4 Statistical Analysis of the Benchmarking

One important objective of our work is to measure the advances of performance over time of a changing benchmarking. In RoboCup@Home a two-steps methodology for benchmarking analysis was adopted (described in details in [8]) based on the definition of a set of desired functional skills, a set of tests executed by participating teams, the definition of a score system that allow to relate score points to skills in the tests, and the evaluation of the amount of score available and actually gained by the best teams for each functionality in each test. In this way, it is possoble to evaluate the average increase of performance in the given skills over years even when changing the tests, in order to make them closer to real world applications.

The functional abilities that have been considered in the competitions so far are the following.

- *Navigation*, the ability of path-planning and safely navigating to a specific target position in the environment, avoiding (dynamic) obstacles.
- *Mapping*, the ability of autonomously building a representation of a partially known or unknown environment on-line.
- *Person Recognition*, the ability of detecting and recognizing a person.
- *Person Tracking*, the ability of tracking the position of a person over time
- *Object Recognition*, the ability of detecting and recognizing (known or unknown) objects in the environment
- *Object Manipulation*, the ability of grasping or moving an object
- *Speech Recognition*, the ability of recognizing and interpreting spoken user commands (speaker dependent and speaker independent)
- *Gesture Recognition*, the ability of recognizing and interpreting human gestures
- *Cognition*, the ability of understanding and reasoning about the world, beyond current perceptions.

As already mentioned in the previous section *Cognition* is a new ability that has been introduced in 2010.

4.1 Analysis of Desired Abilities in the Tests

The first step of our methodology is to define a *functionality-score table* that is used to decide how to distribute the total score in the different desired abilities. This is generated through an iterated process that takes into account both the plans of the Executive and Technical Committees and the feedback from the teams.

Table 1. Functionality-score tables since 2008

Ability	2008	2009	2010
Navigation	40%	33%	22%
Mapping	3%	3%	9%
Person Recognition	10%	12%	12.5%
Person Tracking	6%	4%	3%
Object Recognition	13%	17%	7.5%
Object Manipulation	18%	17%	14%
Speech Recognition	7%	8%	15%
Gesture Recognition	3%	6%	3.5%
Cognition	-	-	13%

Table 1 shows the percentage of score for each functionality over time. Navigation is the most dominant ability, because the competition involves mobile robots. However, since the quality of navigation is increasing and reached very good levels, during the years the impact on the total score is progressively reduced, thus leading the teams to focus on other functionalities. Mapping plays a more limited role, since the environment does not change in a significant way during the competition. Tests are performed outside the arena (e.g., in a real supermarket) and the ability of on-line mapping will be tested in a very realistic environment. Person recognition and tracking are also fundamental abilities. While tracking is easy (thus the score is decreasing over time), recognition is more difficult (score is increasing). In this way research on recognition is stimulated more than, for example, tracking. Object recognition and manipulation also play an important role, however these abilities are in general much more difficult than those related to people. In particular, object manipulation has reached good results only in the last year (see below). Therefore the score for these two functionalities is varying according to the difficulty of the tasks and to the actual accomplishments of the teams. Speech and gesture recognition are needed to implement effective human-robot interaction behaviors. Although in many tests the use speech or gesture is left to the teams, speech is largely preferred. In 2009, we tried to stimulate gesture recognition by increasing the available score, but it did not work (see below) since teams continued to use speech. In 2010 we have devised a test in which the robot has to understand human commands from a distance of 3 meters in a noisy environment, where it is expected that speech will be very challenging and gesture may be even more convenient. Finally, as already mentioned, we decided to assign an important percentage of the score to measure cognitive abilities, in order to focus more on general purpose and robust solutions.

4.2 Analysis of Team Performance

The second step of the methodology is to analyze the actual performance of the teams in these abilities during the competitions. Here we present the results of RoboCup@Home 2008 and 2009.

Table 2. Achieved scores for the desired abilities. 'max' means maximum score achieved and 'av' stands for average score achieved

Ability	2008 max	2008 av	2009 max	2009 av
Navigation	40%	25%	47%	40%
Mapping	100%	44%	100%	92%
Person Recognition	32%	15%	69%	37%
Person Tracking	100%	81%	100%	69%
Object Recognition	29%	8%	39%	23%
Object Manipulation	3%	1%	48%	23%
Speech Recognition	87%	37%	89%	71%
Gesture Recognition	0%	0%	0%	0%
Average	**41%**	**21%**	**61.5%**	**44.4%**

Table 2 presents the percentage of the available scores actually gained by the teams during 2008 and 2009 competitions, related to each of the desired abilities. The second and fourth columns show the best result obtained by some team, while the third and fifth ones contain the average of the results of the five finalist teams.

This table allows for many considerations, such as: 1) which abilities have been mostly successfully implemented by the teams; 2) how difficult are the tests with respect to such abilities; 3) which tests and abilities need to be changed in order to guide future development into desired directions.

From this table, two important aspects are evident. First, the general increase of performance in all the functionalities from 2008 to 2009. Second, the functionalities that have been almost completely solved, and those that are not.

In fact, teams obtained good results in navigation, mapping, person tracking and speech recognition (average above 50%, except for navigation). Notice that the reason for a low percentage score in navigation is not related to inabilities of the teams, but to the fact that part of the navigation score was available only after some other task was achieved. The good results for speech recognition is very relevant since the competition environment is much more challenging than a typical service or domestic application due to a large amount of people and a lot of background noise. On the other hand, the results for mapping and person tracking are due to the fact that they were not applied in a difficult environment.

As already mentioned, this changed in 2010 since new tests in which person tracking and mapping are important will be executed in real environments (a shopping mall and a supermarket). Person recognition performance is acceptable and thus in 2010 we increased a little bit the difficulty of this functionality during the tests. For example, more unknown people will be present during the tests, passing between the robot and the leader during person following, in order to test robustness of the developed methods. Object manipulation had in 2009 the highest increase of performance. Although more robust solutions have been developed by the teams, there is still some work to be done. Therefore for 2010 we did not increase the difficulties of object manipulation in the tests. Also object recognition is reaching good performance and will not become more difficult. Finally, we recognize a problem in gesture recognition which has not been implemented by teams. In fact, only one team (not within the finalists) implemented effective gesture recognition techniques. As already mentioned, the increase of available score was not sufficient to motivate teams to work in this direction. So for 2010 we have created situations where speech is likely to fail and gesture may be the only practical solution to solve the problem.

Summarizing, by analyzing the results of team performance it is possible to decide about future development of the benchmarks. Possible adjustments are:

- Increasing the difficulty if the average performance is high
- Merging of abilities into high-level skills, more realistic tasks
- Keeping or even decreasing difficulty if the observed performance is not satisfying
- Introducing new abilities and tests

As the integration of abilities will play an increasingly important role for future general purpose home robots, this aspect should be especially considered in the future competitions.

Other important parameters to assess the success of a benchmark are the number of participating research groups (teams) and the general increase of performance over the years. Obviously, it is difficult to determine such measures in a quantitative way: the constant evolution of the competition with its iterative modification of the rules and of the scores do not allow a direct comparison. However, it is possible to define some metrics of general increase of performance. They are based on the capability of a team to gain score in multiple tests, thus showing the effectiveness not only in implementing the single functionalities, but also in integrating them in a working system, as well as in the realization of a flexible and modular architecture that allow for executing different tasks.

In Table 3, the number of teams participating to the international competition is shown in the first row. The league received a lot of interest since the beginning (2006), then we registered a general increase. 27 teams (out of 32 requests) have been qualified and are allowed to participate for RoboCup@Home 2010. The second row contains the percentage of successful tests, i.e., tests where some score greater than zero was achieved, showing a significant and constant increase from 17% in 2006 to 83% in 2009. The third row shows the increase in the total number of tests executed by all the teams during the competition. The

Table 3. Measures indicating general increase of performance

Measure	2006	2007	2008	2009
Number of teams	12	11	14	18
Percentage of succ. tests	17%	36%	59%	83%
Total amount of tests	66	76	86	127
Avg. succ. tests p. team	1.0	2.5	4.9	7.3

execution of over 100 tests in 2009 confirms the significance of the statistical analysis we are performing. Finally, the fourth row contains the average number of successful tests for each team. This is a very important measure, since the enormous increase from 1.0 tests in 2006 to 7.3 in 2009 is a strong indication for an average increase in robot abilities and in overall system integration. A team successfully participating in an average of 7 tests (that are quite different each other) demonstrates not only effective solutions and implementation of all the desired abilities, but also a flexible integrated system that has important features for real world applications. Notice that in this table all the teams were considered (not only the five finalists).

The results obtained by the analysis reported here clearly show that our methodology of dynamic benchmarking is producing a quick and significant progress in domestic service robotics.

5 Discussion

The benchmarking high-level robot cognition has just started. The correct paradigm has to be established. RoboCup@Home is actively researching this flexible benchmarking by means of the organization of a world-wide effort and measuring the progress over the years. Although the benchmarking of physical capabilities of the robot in dynamic and poorly structured environments is still in development, there should also be a focus on high-level cognitive tasks that the robot has to perform.

The statistical procedures developed in this competition are useful to start the discussion on the topic of high-level cognitive benchmarking of robots. The methodology can probably be improved and further discussion is needed. RoboCup@Home tries to stimulate not only the benchmarking itself, but also the meta-process about how to benchmark.

The results described in this paper provide evidence that dynamic benchmarking is a viable approach. It can probably be used in many more real-world settings with high levels of uncertainty.

Acknowledgments. The authors gratefully acknowledge the RoboCup Federation for their support over the past five years. We also acknowledge the effort of Thomas Wisspeintner in setting up the league and the help of Stefan Schiffer in making it progress.

References

1. Doostdar, M., Schiffer, S., Lakemeyer, G.: A Robust Speech Recognition System for Service-robotics Applications. In: Iocchi, L., Matsubara, H., Weitzenfeld, A., Zhou, C. (eds.) RoboCup 2008. LNCS, vol. 5399, pp. 1–12. Springer, Heidelberg (2009)
2. Feil-Seifer, D., Skinner, K., Mataric, M.J.: Benchmarks for evaluating socially assistive robotics. Interaction Studies 8(3), 423–439 (2007)
3. Kitano, H., Asada, M., Kuniyoshi, Y., Noda, I., Osawa, E.: Robocup: The robot world cup initiative. In: AGENTS 1997: Proceedings of the First International Conference on Autonomous Agents, New York, NY, USA, pp. 340–347. ACM (1997)
4. Kitano, H., Asada, M., Kuniyoshi, Y., Noda, I., Osawa, E., Matsubara, H.: RoboCup: A Challenge Problem for AI. AI Magazine 18(1), 73–85 (1997)
5. van der Zant, T., Wisspeintner, T.: RoboCup X: A Proposal for a New League Where RoboCup Goes Real World. In: Bredenfeld, A., Jacoff, A., Noda, I., Takahashi, Y. (eds.) RoboCup 2005. LNCS, vol. 4020, pp. 166–172. Springer, Heidelberg (2006)
6. van der Zant, T., Wisspeintner, T.: RoboCup@Home: Creating and Benchmarking Tomorrows Service Robot Applications. In: Robotic Soccer, pp. 521–528. I-Tech Education and Publishing (2007)
7. Wilcox, T., Baillargeon, R.: Object individuation in infancy: The use of featural information in reasoning about occlusion events. Cognitive Psychology 37(2), 97–155 (1998)
8. Wisspeintner, T., van der Zant, T., Iocchi, L., Schiffer, S.: Robocup@home: Scientific Competition and Benchmarking for Domestic Service Robots. Interaction Studies. Special Issue on Robots in the Wild 10(3), 392–426 (2009)

Requirements and Platforms for Social Agents That Alarm and Support Elderly Living Alone

Marleen E. Spiekman[1,2], Pascal Haazebroek[2], and Mark A. Neerincx[1,3]

[1] TNO, Department of Perceptual and Cognitive Systems, Soesterberg, Netherlands
{marleen.spiekman,mark.neerincx}@tno.nl
[2] Leiden University, Cognitive Psychology Unit, Leiden, Netherlands
phaazebroek@fsw.leidenuniv.nl
[3] Delft University of Technology, Man-Machine Interaction Group, Delft, Netherlands

Abstract. Social embodied agents may mitigate moments of apathy and confusion that older adults can experience at home. Based on a literature study, use cases, requirements and claims were specified. In an experiment with 29 older adults (aged 70+), it was studied to what extent a virtual agent and three robots (i.e., the Nao, iCat and Nabaztag) provide a platform to support these use cases, requirements and claims. Participants seemed to evaluate the agents mainly in terms of three generic components: the perceived level of realism, intellectuality and friendliness. A more serious and agreeable appearance improved the appreciation of the agent's actions. Especially facial realism appeared to be important for trust, social presence, perceived sociability and perceived enjoyment.

Keywords: Mild dementia, elderly, assistive technology, social agent.

1 Introduction

Worldwide the population is growing older. Although there are large differences among older people, this group is generally subject to physical and cognitive decline, increasing the demand on healthcare services [1]. A step further than normal cognitive decline is mild dementia. Intelligent technology can support elderly to cope with the consequences of cognitive decline and mild dementia and help people to live longer independently [2]. Social embodied agents are particularly promising for this purpose, due to their easiness of use and multiple possibilities [3]. A main issue is how social agents can support elderly people to live longer independently and particularly how social agents can support mild demented elderly living alone.

Different platforms for social agents provide these agents with different user interfaces, resulting in different appearances and functionality for non-verbal communication. Examples of social agents that possibly can take a supporting role for elderly are Ashley, developed by TNO [4] and iCat and Nao, which both have been used in studies with elderly [5-6]. Specific advantages of these platforms are that their interaction is highly intuitive and personalized. A simpler platform used in studies with elderly is Nabaztag [7], which is less sophisticated but commercially available and

B. Mutlu et al. (Eds.): ICSR 2011, LNAI 7072, pp. 226–235, 2011.

cheap, and therefore attractive for large scale use. Table 1 gives the main differences among the agents, as they were rated by the experimenters.

This paper describes our first experiment with these four agents, which we performed with 29 elderly (aged 70+). This study is our first step in the iterative design process of social agents to find out how social agents can help senior adults with mild dementia to live alone.

Table 1. The four platforms and their main differences (as rated by the experimenters)

	 **Ashley**	 **iCat**	 **Nao**	 **Nabaztag**
Human-like	++	+	0	--
Animal-like	--	--	++	++
Machine-like	-	+	0	0
Toy-like	--	0	+	++
Body realism	++	+	-	--
Face-realism	++	++	-	--

2 Background

2.1 Needs of Demented Elderly Living Alone and Assistive Technology

Due to the development of their disease, people with dementia living alone have additional needs, which can't all be met by caregivers, neighbors and friends [8]. Several studies show unmet needs in the domains of memory and daytime activities, general and personalized information, social contact and company, and health monitoring and personal safety e.g. [8-9]. Many technological developments translate these needs into assistive technology [10].

In the development of assistive technology the interaction between the technology and the elderly is an important aspect that deserves more attention. Currently assistive innovations often use sound, audio messages, text and/or pictures on a screen to communicate with users, e.g. [10-11]. Social agents can be an added value here, since they are typically useful for intuitive and persuasive interaction, adapting to a large variety of users and remaining attractive to users on a long term basis [12]. In addition, social agents can be used as sensor by asking questions to the user.

2.2 Social Robots for Demented Elderly

Heerink [3] gives an overview of social robots that have specifically been developed for older adults, in which we distinguishes between companion type robots and service type robots. Whereas companion type robots like Paro are especially used with demented elderly, the studies on service type social agents don't have a specialized focus on senior adults with dementia. Furthermore, most studies use only one platform. So we conclude that there isn't much knowledge on platforms for social agents supporting senior adults with mild dementia to live alone.

3 Research Approach

3.1 Situated Cognitive Engineering (SCE)

We use the Situated Cognitive Engineering (SCE) approach [13] to develop a social agent to alarm and support people with mild dementia living alone. SCE is an interactive incremental design process that combines operational demands (needs and context of people with mild dementia living alone) with user centered design (human factors knowledge) and technological centered design (the envisioned technology which can make the development possible). Integration of these aspects leads to the core functions and use cases to which the design needs to answer. Claims justify the requirements and reveal a possible trade-off between advantages and disadvantages of these requirements. Evaluating the claims leads to refinement of the requirements of the design. This design process is schematized in fig. 1. SCE is used because it is a systematic method to acquire, define and validate design knowledge and step by step adjust and refine the functionality of the design based on this increasing knowledge. [14]

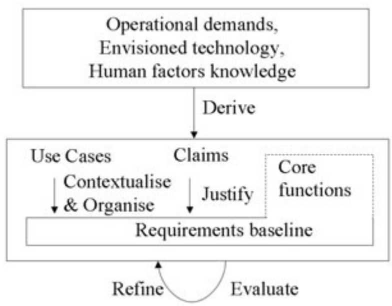


Fig. 1. Situated Cognitive Engineering (SCE) methodology

3.2 First Requirements, Claims and Use Cases

Since demented elderly are a vulnerable group and the development of social agents for this group is in its infancy, we anticipated that involving mild demented elderly in this first phase would be too burdensome for them. Hence, we decided to perform the first development steps together with healthy elderly.

Based on a literature study, we specified a set of use cases describing the general behavior requirements in their context (see table 2 for a summary), and we formulated a series of requirements, see table 3. The requirements are accompanied by corresponding confirming and disaffirming claims. If and which claims hold is subject to this and future studies. The assumption of this design approach is that agents that meet the requirements will perform better. What is good performance is determined by the use cases. In our case the agents should be able to put a person at ease, give trustworthy advise and is listened to if needed. This leads to the following research questions (Q) and hypotheses (H) for this study:

- The agents in the study differ on various aspects (see table 1). How do elderly (aged 70+) evaluate them in terms of basic and personality characteristics and perceived performance (Q1) and what aspects determine their evaluation (Q2).
- Agents that are perceived as showing more non-verbal communication and empathy will be rated higher in terms of trust, social presence, perceived sociability and perceived enjoyment (H1).
- The perceived performance of the agent will be higher if the agent is perceived as more dominant, extravert, serious, agreeable, and/or intelligent at appropriate times (H2).

Table 2. Summary of the three use cases used in our experiment [15]

Use case 1: Drink	Use case 2: Fall-check	Use case 3: Wandering
Demented elderly sometimes forget to drink. The agent helps the user to remember, and encourages him/her to drink more. The user drinks more and remains more in control.	Demented elderly are prone to fall more quickly than others. Sensors can detect if the person has fallen, but false alarms do occur. The agent checks if help is needed and alarms a third person if necessary. The user is helped if needed and remains in control over the situation as far as possible.	Some demented elderly wander at night. Since this is potentially dangerous, the agent tries to prevent the user from leaving the house unwontedly. The user is called back inside and put at ease.

Table 3. First requirements and claims [3], [5], [12], [16-18]

Requirements	**Claims (positive)**	**Claims (negative)**
The agent should show non-verbal communication and empathy	Appropriate non-verbal communication and empathy improves trust, perceived sociability, social presence and enjoyment.	If the non-verbal communication isn't congruent with the context, it can give a feeling of unnaturalness and it can be distractible (especially for senior/demented adults with decreased working memory capacity).
The agent should look and behave dominant, assertive, serious, friendly and/or intelligent.	Dominant, assertive, serious, agreeable and/or intelligent looks and behavior are congruent with part of the tasks and that will increase compliance.	People will feel uncomfortable if personality and situational context aren't congruent; Decreases enjoyment, which can influences long-term relationship / compliance on the long term.

4 Experiment

4.1 Method

Twenty-nine people participated in the experiment (male: N=20, female: N=9), with age ranging from 70 to 85 years (mean = 75,9; standard deviation = 4,91). We confronted the participants in a one-on-one communication with the four agents. The communication was done in a Wizard of Oz setting, which means that the participants thought they were interacting with an autonomous agent, while in reality the experimenter operated the agents from an adjacent room using the same fixed scripts for all agents, based on the same texts, voice and intonation. The scripts also included emotion and movements, depending on the functionality of the agents (Ashley: lips and balance; Nao: head, arms and body; iCat: head, lips, eyes and eyebrows; Nabaztag: ears and led lights). The participants all communicated with all four agents in counterbalanced order and in the three use cases in the same order.

The participants were asked to enter into the role of a demented elderly living alone. After being explained the general procedure and the first use case, the participant was left alone to communicate with the first agent. The agent always initiated the conversation. After the first conversation they filled out a short questionnaire about the perceived performance of the agent (see table 5 first column, assessed by a 7 point Likert scale), this procedure repeated itself for use case two and three. Then the participant filled out a longer questionnaire, evaluating the agent on basis characteristics (human-, animal-, machine- and toylikeliness, intelligence, seriousness, dominance, adorability, dynamic-ness and non-verbal communication; each assessed by a 7 point Likert scale) and personality (extraversion and agreeableness; each based on 3 subquestions with a 9 point Likert scale each). This procedure was repeated for all agents.

Then the participants were asked to perform a paired comparison between all combinations of the agents, measuring trust, social presences, perceived sociability and perceived enjoyment. The experiment ended with a short interview in which the participants were asked which agent they would want to have at home when they would live alone with mild dementia and how they evaluated each agent in general.

4.2 Results

Comparing Agents on Characteristics

ANOVA analyses were used to compare the characteristics of the agents. On average, the agents only differed significantly on how human-like, animal-like, machine-like and toy-like the elderly perceived them (see table 4). The average scores generally correspond to our own ratings in Table 1, in which we judged the agents on these characteristics upfront. No differences were found on the perceived performance and on the other characteristics which were measured (agreeable, extravert, dominant, intelligent, serious, adorable, dynamic and non-verbal communication).

Table 4. Average scores and ANOVA F and p values of four of the basic characteristics (incl. standard deviations) for the four agents, measured with a 7 point Likert scale (1=low, 7=high)

	Ashley	Nao	iCat	Nabaztag	F(3,112)	p
Human-like	4,9 (1,93)	3,7 (1,89)	3,3 (1,93)	3,2 (1,91)	4,98	< .01
Animal-like	1,2 (0,44)	1,6 (1,08)	2,2 (1,78)	3,8 (2,44)	14,41	< .001
Machine-like	3,2 (2,13)	5,4 (1,68)	5,0 (1,85)	4,6 (2,13)	6,93	< .001
Toy-like	2,7 (1,99)	4,7 (2,07)	5,4 (1,35)	5,6 (1,68)	16,24	< .001

Predicting Performance by Characteristics

In order to analyze whether the perceived performance was positively influenced by any of the agent characteristics (H2), we performed stepwise multiple regression analyses with the performance measures as dependent variables and the basic and personality characteristics as possible predictors. The predictor variables ‘agreeableness’ and ‘seriousness’ showed moderate standardized regression coefficients (β) for most of the performance measures (see table 5), which is a measure of how strong these predictors influence the criterion variable: ‘agreeableness’ and ‘seriousness’ appear to be moderate predictors for most of the performance measures.

Clustering Characteristics on Participants Evaluation

One of the research questions was ‘what aspects determine the senior adults’ evaluation of the agents’ (Q2). We used a Principle Component Analysis (PCA) to analyze whether the characteristics we measured in the experiment might be clustered together in a reduced amount of categories of characteristics that measure generally the same and that might indicate these general evaluation criteria. These categories might represent different generic underlying components on which participants evaluate the agents. The PCA resulted in 3 groups of components. The factor loadings on the components for the characteristics are shown in table 6.

Table 5. Multiple regression results (R^2 = the proportion of the variance in the criterion variable which is accounted for by our model, ΔR^2 = the change R^2 per step, β = standardized regression coefficient, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$, non significant variables are not shown)

Dependent variables	R^2	β of predictive variables: Serious	Agreeable	Extravert	Adorable	Dominant	Machine-like
Extent in which agents put you at ease	$,38^{***}$	$,35^{***}$ (Step 1 $R^2 = .22$)	$,46^{***}$ (Step 2 $\Delta R^2 = .11$)		$-,19^{*}$ (Step 3 $\Delta R^2 = .02$)	$-,16^{*}$ Step 4 $\Delta R^2 = .02$)	
Extent in which you’d follow agent’s advise	$,50^{***}$	$,52^{***}$ (Step 1 $R^2 = .36$)	$,42^{***}$ (Step 2 $\Delta R^2 = .07$)		$-,23^{**}$ (Step 3 $\Delta R^2 = .04$)		$,17^{*}$ (Step 4 $\Delta R^2 = .03$)
Extent in which you’d do what agent asks	$,45^{***}$	$,49^{***}$ (Step 1 $R^2 = .32$	$,17^{*}$ (Step 4 $\Delta R^2 = .02$)	$,24^{**}$ (Step 2 $\Delta R^2 = .09$)			$,17^{*}$ (Step 3 $\Delta R^2 = .02$)

Table 6. PCA results: The factor loadings on the components for the characteristics (note that factor loadings below 0,45 aren’t shown in the table; Comp. = Component)

Characteristic:	Comp. 1	Characteristic:	Comp. 2	Characteristic:	Comp. 3
Toy-like	-,75	Serious	,79	Adorable	,82
Human-like	,73	Intelligent	,77	Agreeable	,67
Animal-like	-,64	Extravert	,64	Non-verbal communicative	,53
Dominant	,63				
Machine-like	-,56			Dynamic	,52

The first component scores highly positive on human-like and dominance and highly negative on toy-, animal-, and machine-like, therefore we interpret this component is perceived level of realism. The second component relates to seriousness, intelligence and extraversion. We interpret this as the intellectual, sensible or rational component of the agents. The third component scores high on adorability, agreeableness, non-verbal communication and dynamic behaviour. We interpret this as friendliness or likability and links to empathic behaviour.

Comparing Agents on Participant Preferences

The participants were asked to compare and choose pair wise between the agents on four aspects: trust, social presence, perceived sociability and perceived enjoyment. Since there was a strong correlation among these four aspects, we interpreted the average score on these aspects as the general preference of the participants for the specific agent. For the overall results see fig. 2. Although no statistical judgment can be made from these data, Nao and Nabaztag are less preferred and iCat and Ashley are most preferred.

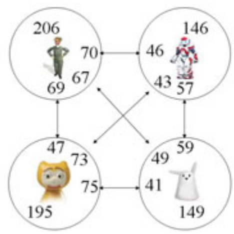


Fig. 2. Paired comparison results. The values at the end of the arrows show the amount of times the agent is chosen over the one at the other end of the arrow (sum of the 4 paired comparison questions). The single value is the total amount of times the agent is chosen.

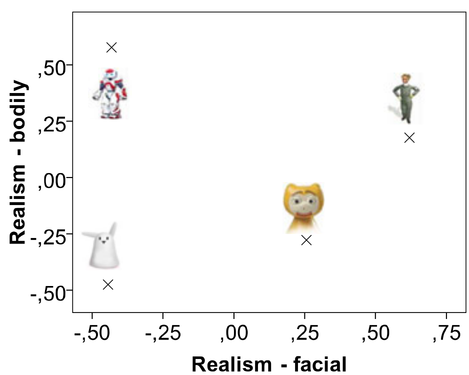


Fig. 3. Multidimensional scaling plot of the general preference for the four agents. (Note that the axes were labeled afterwards, based on the results in the graph.)

To gain more insight to what extent the agents resemble each other a multidimensional scaling technique (MDS) was used to analyze these paired comparisons of the general preference, see fig. 3. The plot gives an insight in the participant's perception of the agents. The x-axis and the y-axis in the plot represent two (perceived) dimensions or characteristics on which the participants evaluate the similarity of the agents (distances in the plot represent the level of similarity: dots that are close together represent items that are judged as relatively similar).

The scores on the x- and y-axes correspond with the pre-ratings of 'facial realism' and 'bodily realism' of the agents by the experimenter (see Table 1. Note that we didn't ask the participants to score facial and bodily realism directly). We assume that the two perceived characteristics which the participants use to compare the agents might be these two aspects of realism of the agent: facial realism and bodily realism.

On the question 'Which agent would you like to have at home', 11 participants chose Nabaztag, 10 chose Ashley, 4 chose Nao and also 4 chose iCat. People found Ashley the most realistic and Nabaztag adorable, simple and unobtrusive.

5 Conclusions and Discussion

On the research question how elderly evaluated the agents and what aspects determined their evaluation, we conclude that in terms of perceived basic and personality characteristics and perceived performance elderly evaluate the four agents as rather similar, only judging Ashley as more human-like, less machine-like and toy-like and Nabaztag as more animal-like than most of the others (Q1). In a Principal Component Analyses we found three components which might determine their evaluation, which we interpreted as the perceived level of realism, intellectuality and friendliness (Q2). Despite the differences in appearance among the agents, differences among the agents' characteristics were only perceived on the level of realism, while participants evaluated the level of friendliness and intellectuality as not significantly different. However, the results show that these differences on perceived realism didn't influence the perceived performance of the agents. The agents didn't differ on what they said and how they said this (content, type and tone of voice). Perhaps these aspects are more important to determine the perceived level of friendliness and intellectuality of the agents than mere looks. And the level of friendliness and intellectuality might on their turn be more important for perceived performance than looks. Several studies indeed imply an important role for content, type and tone of voice: The voice of an agent is shown to affect the way people perceive a robot [20] and in their study on people's mental models of social agents Kiesler & Goetz [21] found that dialogue had a larger effect on the perception of the robot's personality than appearance.

Although we found only a limited amount of differences among the agents' characteristics and performances, this doesn't mean that the different appearances of the agents didn't influence the perceived performance at all. We hypothesised that the perceived performance of an agent will be higher if the agent is perceived as more dominant, extravert, serious, agreeable and/or intelligent at appropriate times (H2). Indeed we found that agreeableness and seriousness are moderate predictors for most of the performance measures: people who assess an agent as more agreeable and/or serious, based on its appearance, also perceive the agent as better performing. This finding corresponds with the finding of Goetz et al. [18], who found that a robot induced more compliance on a task if the appearance and behaviour of the robot matched the task requirements.

Although we didn't find a direct link between perceived realism and performance, the results do suggest a link between realism and user preference. We performed analyses of the paired comparison between all paired combinations of agents on trust, social presence, perceived sociability and perceived enjoyment. This analysis indicate that the participants evaluated the agents on two different aspects of realism: facial realism and bodily realism. Of these two the user preference appears to be influenced by facial realism mainly: the more facial realism an agent has, the more it is preferred on the mentioned aspects. The hypothesis that non-verbal communication and empathy will be preferred more in terms of trust, social presence, perceived sociability and perceived enjoyment (H1) isn't proven by this experiment, but it is plausible that facial realism is a vital link in this.

Surprisingly, on the question which agent one prefers to have at home, the most unrealistic agent (Nabaztag) scored as high as the most realistic agent (Ashley). So presumably, realism isn't the key characteristic for an agent in the home setting. Arguments for choosing Nabaztag were that she is adorable, small, simple and unobtrusive. It is found in literature that in a home setting older people prefer small robots [20]. Broadbent et al. [20] found that older people are less comfortable with new technology and computers than younger people and older people also have less trust in robots. Due to the anxiety older people feel towards robots, they say they prefer robots that 'appear less threatening, are small, have female voice, move slow, are less autonomous, have a serious aspect and have a single colour'. Another explanation might be that even an image of a pair of eyes make people feel watched and change their behaviour [19]. Maybe some senior adults prefer Nabaztag because being less realistic makes her less unobtrusive and gives people less the feeling of being watched in their home environment. Different user characteristics might play a role here. Finally, the preference for both Ashley and Nabaztag might be explained by the alignment of the facial and bodily realism in these agents: perhaps people don't mind the level of realism as long as it is consistent for all aspects of the robot.

On this topic, we conclude that realistic facial features are an important characteristic of a social agent in the role of supporting elderly at home, but that an agent in the living room should be small and not be too disturbing and too present, at least for some. A challenge will be to combine realism with low obtrusiveness. Perhaps users should have a choice among platforms, which might call for flexibility in the link between font-end platforms and back-end artificial intelligence.

We performed our experiment with healthy adults aged 70+. The development of dementia is a gradual process [15] and we don't expect people with mild dementia to perceive social agents as substantially different as our healthy participants did. Of course when people become more and more confused this changes. Therefore the next step will be to evaluate the performance of social agents with mild demented elderly in their home environment and also for longer periods.

References

1. United Nations: World Population Ageing. Summary. United Nations Publications, New York (2007)
2. Pollack, M.A.: Intelligent technology for an aging population: the use of AI to assist elders with cognitive impairment. AI Magazine 26, 9–24 (2005)

3. Heerink, M.: Assessing acceptance of assistive social robots by aging adults. PhD thesis, University of Amsterdam (2010)
4. Nauts, P., van Doesburg, W., Krahmer, E., Cremers, A.: Taking turns in flying with a virtual wingman. In: Proc. of HCI International 2011, Hilton Orlando Bonnet Creek, Orlando, Florida, USA, July 9-14 (accepted, 2011)
5. Looije, R., Neerincx, M.A., Cnossen, F.: Persuasive robotic assistant for health self-management of older adults: Design and evaluation of social behaviors. International Journal of Human-Computer Studies 68, 386–397 (2010)
6. KSERA, EU project website, `http://ksera.ieis.tue.nl/`
7. Klamer, T., Allouch, S.B.: Zoomorphic robots used by elderly people at home. In: Proc. of the 27th Int. Conf. on Human Factors in Computing Systems. ACM International Conference on Human Factors in Computing Systems. ACM Press, Atlanta (2010)
8. Lauriks, S., Reinersmann, A., van der Roest, H.G., Meiland, F., Davies, R., Moelaert, F., Mulvenna, M.D., Nugent, C.D., Dröes, R.-M.: Review of ICT-based services for identified unmet needs in people with dementia. Ageing Research Reviews 6, 223–246 (2007)
9. van der Roest, H.G., Meiland, F.J.M., Comijs, H.C., Derksen, E., Jansen, A.P.D., van Hout, H.P.J., et al.: What do community-dwelling people with dementia need? A survey of those who are known to care and welfare services. International Psychiatrics 21, 949–965 (2009)
10. Mulvenna, M.D., Nugent, C.D. (eds.): Supporting people with dementia using pervasive health technologies. Springer, London (2010)
11. eHome, `http://deutsch.ceit.at/ceit-raltec/projekte/aal—ehome`
12. Fong, T., Nourbakhah, I., Dautenhahn, K.: A survey of socially interactive robots. Robotics and Autonomous Systems 42, 143–166 (2003)
13. Neerincx, M.A., Lindenberg, J.: Situated cognitive engineering for complex task environments. In: Schraagen, J.M.C., Militello, L., Ormerod, T., Lipshitz, R. (eds.) Naturalistic Decision Making and Macrocognition, pp. 373–390. Ashgate Publishing Limited, Aldershot (2008)
14. Westera, M., Boschloo, J., van Diggelen, J., Koelewijn, L.S., Neerincx, M.A., Smets, N.J.J.M.: Employing use-cases for piecewise evaluation of requirements and claims. In: Proc. of ECCE, pp. 279–286. Delft University of Technology, Netherlands (2010)
15. Alzheimer Nederland: `http://www.alzheimer-nederland.nl`
16. Cramer, H.S.M.: People's responses to autonomous and adaptive systems. PhD thesis, University of Amsterdam (2010)
17. Bickmore, T., Schulman, D., Yin, L.: Maintaining engagement in long-term interventions with relational agents. Applied Artificial Intelligence 24, 648–666 (2010)
18. Goetz, J., Kiesler, S., Powers, A.: Matching robot appearance and behaviour to tasks to improve human-robot cooperation. In: Proc. of the 12th IEEE Int. Workshop on Robot and Human Interactive Communication, Lisbon, Portugal. Roman, Millbrae, CA (2003)
19. Bateson, M., Nettle, D., Roberts, G.: Cues of being watched enhance cooperation in a real-world setting. Biology Letters 2, 412–414 (2006)
20. Broadbent, E., Stafford, R., MacDonald, B.: Acceptance of healthcare robots for the older population: Review and future directions. International Journal of Social Robotics 1, 319–330 (2009)
21. Kiesler, S., Goetz, J.: Mental models and cooperation with robotic assistants. In: Proceedings of Conference on Human Factors in Computing Systems. ACM, New York (2002)

Author Index