Mixed Reality and Human-Robot Interaction

International Series on
INTELLIGENT SYSTEMS, CONTROL, AND AUTOMATION:
SCIENCE AND ENGINEERING

VOLUME 47

For other titles published in this series, go to
www.springer.com/series/6259

Xiangyu Wang

Mixed Reality and Human-Robot Interaction

Springer

Xiangyu Wang
Senior Lecturer
Leader of Virtual Technology Group
Construction Management and Property
Faculty of Built Environment
The University of New South Wales
Sydney, NSW
Australia
Email: xiangyu.wang@unsw.edu.au

ISBN 978-94-007-0581-4 e-ISBN 978-94-007-0582-1
DOI 10.1007/978-94-007-0582-1
Springer Dordrecht Heidelberg London New York

Typesetting & Cover design: Scientific Publishing Services Pvt. Ltd., Chennai, India

Printed on acid-free paper

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

Preface

In the recent past, Mixed Reality (MR) technologies play an increasing role in Human-Robot Interactions (HRI) such as telerobotics. The visual combination of digital contents with real working spaces creates a simulated environment that is set out to enhance these interactions. A variety of researches explored the possibilities of Mixed Reality and the area of human-robot interaction. From a thorough review of competitive books in both areas, it was found that there has not been a collected publication that focuses on integration of MR application into Human-Robot Interaction in the context of all kinds of engineering disciplines, although there are only 20-30 noted researchers in the world who are now focusing this new, emerging, and cutting-edge interdisciplinary research area. This area is expanding fast from what were observed in the new special sessions/themes/workshops of leading international research conferences. The book addresses and discusses fundamental scientific issues, technical implementations, lab testing, and industrial applications and case studies of Mixed Reality in Human-Robot Interaction. Furthermore, more and more researchers in applying MR in these areas emerge and need a guide to bring the existing state-of-the-art into their awareness and start their own research quickly. Therefore, there is as strong need to have a milestone-like guidance book for following researchers who are interested in this area to catch up the recent progress.

The book is a reference book that not only acts as meta-book in the field that defines and frames Mixed Reality use in Human-Robot Interaction, but also addresses up-coming trends and emerging directions of the field. The target audiences of the book are practitioners, academics, researchers, and graduate students at universities, and industrial research that work with Mixed Reality and Human-robot interaction in various engineering disciplines such as aerospace, mechanical, industrial, manufacturing, construction, civil, and design, and also the disaster research and rescue.

The book addresses a variety of relevant issues in Mixed Reality (MR). Chapters covering the state-of-the-art in MR applications in all areas of human-robot interactions and how they can be applied to influence the human-robot interface design and effectiveness in various engineering disciplines such as aerospace, mechanical, industrial, manufacturing, construction, civil, and design, and also the disaster research and rescue. The results of most recent internationally most renowned inter-disciplinary research projects presenting and discussing application solutions of MR technologies in Human-Robot Interaction. The topics covered by the book include psychological fundamentals in Human-Robot Interaction, innovative concepts of integrating Mixed Reality and Human-Robot

Interaction, the development/implementation of integrating Mixed Reality and Human-Robot Interaction, and evaluation of Mixed Reality-based Human-Robot Interactions.

This book offers a comprehensive reference volume to the state-of-the-art in the area of MR in Human-Robot Interaction. This book is an excellent mix of over 9 leading researcher/experts in multiple disciplines from academia and industry. All authors are experts and/or top researchers in their respective areas and each of the chapters has been rigorously reviewed for intellectual contents by the editorial team to ensure a high quality. This book provides up-to-date insight into the current research topics in this field as well as the latest technological advancements and the best working examples.

To begin, James E Young, Ehud Sharlin, and Takeo Igarash, the terminology of Mixed Reality in the context of robotics, in their chapter *What is Mixed Reality, ANYWay? Considering the Boundaries of mixed reality in the Context of Robots*. They clarified the definition of MR as a concept that considers how the virtual and real worlds can be combined rather than a class of given technology. Further, they posit robots as mixed-reality devices, and present a set of implications and questions for what this implies for MR interaction with robots.

The second chapter *User-Centered HRI: HRI Research Methodology for Designers* by Myungsuk Kim, Kwangmyung Oh, Jeong-Gun Choi, Jinyoung Jung, and Yunkyung Kim, introduces the field of user-centered HRI, which differs from the existing technology-driven approach adopted by HRI researchers in emphasizing the technological improvement of robots. It proposes a basic framework for user-centered HRI research, by considering three main elements of "aesthetic", "operational", and "social" contextuability.

Human-robot interfaces can be challenging and tiresome because of misalignments in the control and view relationships. These mental transformations can increase task difficulty and decrease task performance. Brian P. DeJong, J. Edward Colgate, and Michael A. Peshkin discussed, in *Mental Transformations in Human-Robot Interaction*, how to improve task performance by decreasing the mental transformations in a human-robot interface. It presents a mathematical framework, reviews relevant background, analyzes both single and multiple camera-display interfaces, and presents the implementation of a mentally efficient interface.

Next chapter, by David B. Kaber, Sang-Hwan Kim and Xuezhong Wang, in *Computational Cognitive Modeling of Human-Robot Interaction Using a GOMS Methodology*, presents a computational cognitive modeling aproach to further understand human behavior and strategy in robotic rover control. GOMS (Goals, Operators, Methods, Selection Rules) Language models of rover control were constructed based on a task analysis and observations during human rover control trials.

During the past several years, mobile robots have been applied as an efficient solution to explore inaccessible or dangerous environments. As another application of Mixed Reality concept into the Robotics, the chapter, *A Mixed Reality-based Teleoperation Interface for Mobile Robot* by Xiangyu Wang and Jie

Zhu, introduces a Mixed Reality-based interface that can increase the operator's situational awareness and spatial cognitive skills that are critical to teleorobotics and teleoperation.

The chapter by Iman Mohammad Rezazadehi, Moohammad Firoozabadi, and Xiangyu Wang, *Evaluating the Usability of Virtual Environment by Employing Affective Measures*, explores a new approach that is based on exploring affective status and cues for evaluating the performance and designing quality of virtual environments.

Building intelligent behaviors is an important aspect of developing a robot for use in security monitoring services. The following chapter, *Security Robot Simulator*, by Wei-Han Hung, Peter Liu, and Shih-Chung Jessy Kang, proposes a framework for the simulation of security robots, called the *security robot simulator* (SRS), which is aimed at providing a fully inclusive simulation environment from fundamental physics behaviors to high-level robot scenarios for developers.

The final chapter by K.L. Koay, D.S. Syrdal, K. Dautenhahn, K. Arent, Ł. Małek, and B. Kreczmer, titled *Companion Migration – Initial participants' feedback from A VIDEO-based Prototyping Study*, presents findings from a user study which investigated users' perceptions and their acceptability of a Companion and associated 'personality' which migrated between different embodiments (i.e. avatar and robot) to accomplish its tasks.

Acknowledgements

I express my gratitude to all authors for their enthusiasm to contribute their research as published here. I am also deeply grateful to my external reader Ms. Rui Wang at The University of Sydney, whose expertise and commitment were extraordinary and whose backup support on things both small and large made the process a pleasant one.

Xiangyu Wang

Contents

Companion Migration – Initial Participants' Feedback from a Video-Based Prototyping Study

K.L. Koay[1], D.S. Syrdal[1], K. Dautenhahn[1], K. Arent[2], Ł. Małek[2], and B. Kreczmer[2]

[1] University of Hertfordshire, U.K.
[2] Wrocław University of Technology, Poland

Abstract. This chapter presents findings from a user study which investigated users' perceptions and their acceptability of a Companion and associated 'personality' which migrated between different embodiments (i.e. avatar and robot) to accomplish its tasks. Various issues such as Companion migration decision, Retention of Companion identity in different embodiments, Personalisation of Companion, users' privacy and control over the technology are discussed. Authorisation guidelines for Companions regarding migration, accessing an embodiment and the data stored in the embodiment are proposed and discussed for future design of migration Companion.

Keywords: Migration Companion, Virtual Agent, Human-Robot Interaction, Social Robotics, Robot Companion.

1 Introduction

The goal of much current Human-Robot Interaction and Human-Computer Interaction research is to create an artificial entity that can support its user in their daily activities and in their own social setting. This artificial entity should be able to create social and emotional relationships with its user through long-term interaction by learning and respecting user's preferences (Koay et al., 2007), habits, requirements, and relate to a user in some of the ways that a human might in a natural way (Bickmore and Picard 2005; Dautenhahn 2004; Goodrich and Schultz 2007).

It would be useful if such an artificial entity were not limited to a single embodiment, such as a particular robot or avatar whose abilities are constrained to their fixed embodiments and functionalities within their particular environment (Kidd and Breazeal 2007; Imai et al. 1999; Koay et al. 2009a). Hence it is beneficial to have a single instance of the artificial entity, that a user is familiar with and that can migrate from one embodiment to another as required to achieve particular goals (O'Hare et al., 2003). The *personality* of the artificial entity here is defined as the *persistent features which make it unique and recognizable from the owner's perspective.* For the purpose of this book chapter we call a Companion an artificial entity with a personality (as defined above) and

X. Wang (Ed.): Mixed Reality and Human-Robot Interaction, ISCA 47, pp. 133–151.
springerlink.com

the capability of migrating between different embodiments to support the user's daily activities.

Being able to migrate between different embodiments allows the Companion to continuously provide assistance to the user (Ogawa and Ono, 2008) and can achieve a stronger sense of contextual and situational awareness of its physical and social environment. This helps improve the Companion's understanding of its user, which should lead to improving their relationship, thus contributing to a sense of Companionship for the user (Koay et al., 2009b). Over time the Companion will establish its own unique identity and beliefs and should be able to maintain a relationship with the user that is unique, individual and personal regardless of the Companion's embodiments (MacDorman and Cowley, 2006). This will provide the user with a strong sense of familiarity and trust that will encourage them to interact with the Companion over a long-term period.

The Companion, in the context of our research project, should provide physical task support, utilizing appropriate physical embodiments, and also to act as a cognitive prosthetic, utilizing simple hand-held devices, such as a mobile phone which the user can easily carry. Furthermore it is easier for the user to take their Companion with them outside (e.g. during travel) when it inhabits a hand-held device rather than transporting a larger scale robot. For example, a Companion would be particular useful (but not limited) to help an elderly user or person with special needs to continue independently living in their own home. The Companion could provide physical assistance, such as helping them to get up from their seats, fetching them their medication at night or by acting as a cognitive prosthetic by reminding them about their glossaries and appointments regardless of where they are.

However, the challenges for any successful new technology include not only identifying the technical aspects of its functionalities and usability but also to examine the users' perspectives. In this chapter, we will explore users' perceptions and their acceptability of a Companion migrating between virtual embodiments, namely a computer generated avatar and a robot with a physical embodiment in the real world, to accomplish its tasks. This includes not only the identification of migratable attributes of the Companion, but also identifying key factors which enable the Companion to successfully express and retain its identity across different embodiments.

Naturally, users have concerns with regard to privacy, usability and applicability (Syrdal et al., 2007) and these were also considered in the study. The main aims of the study are to guide the design of Companions that users want to interact with, and to be useful regardless of the embodiments they inhabit, and to be able to maintain their identity and the users' belief that they are still interacting with the same Companion, regardless of the embodiments it occupies.

2 Research Questions and Methodologies

In order to create a specification of requirements and design guidelines for a migratable Companion, the following research question will need to be addressed:

From a user's perspective, how does he perceive Companions changing embodiments, Companion's retention of identity, user's privacy, context of use and the functionality of migrating Companions?

These issues are vital because if the users have no concept of their Companion being able to migrate between different embodiments, they may interpret the process of migration as a form of communication or collaboration between two artificial entities (Koay et al., 2009b). Therefore, it is important to understand from a user's perspective how a Companion can convey the impression of Companion migration as an aspect of technology where the Companion is able to move between different embodiments as required.

The concept of Companion migration in this chapter is based on a Soul and Shell terminology where the Companion functions as a "SOUL", while other physical devices (i.e. robots and avatars) function as a "SHELL". Only one instance of the Companion exists and is able to migrate between different shells (LIREC Deliverable D8.1, 2009).

2.1 Video-Based Methodology

To reach larger number of participants and to have greater control over a standardised experimental procedure, a video-based methodology (Woods et al. 2006; Newell et al. 2006) was used for this study. Other advantages are that it avoids having to train each individual participant how to interact with the technology, and having technical malfunctions influence the experiences of some participants and biasing their feedback. The video-based methodology was also chosen as it allows participants to visualize the concept of Companion migration and possible application through controlled video narratives. This is a particularly effective method for collecting participants' first impressions and opinions of the technology, without dwelling on any technical aspects that may occur when running actual hands-on trials. It was shown in previous studies to be particular effective for prototyping and collecting participants views without introducing bias (Woods et al. 2006; Walters et al. 2008).

2.2 Back Translation of Questionnaire

The questionnaire was initially created in English and then translated into Polish. In order to gauge the accuracy of this translation, this was in turn translated back into English. This is known as back translation and is an accepted method of assessing the reliability of questionnaire items across languages (Chapman and Carter, 1979). The participants had the option of responding in English or Polish. For the purpose of this analysis, Polish responses were translated into English by Polish native speakers.

2.3 Context of Migration

Being able to change its embodiment allows the Companion to present itself as a consistent presence and therefore should be perceived over time as reliable and trustworthy to its user. However this impression depends strongly on the

Companion context of migration, such as when the Companion should change its embodiments, which embodiments it should migrate to, and who should have the final say on the selection of a particular embodiment. Our approach for investigating this issue is to implement the Companion based on the following two hypotheses:

H1 - Companion should not only have contextual and task awareness, but also a strong sense of the capabilities of the various embodiments it can migrate to in order to be useful and continue assisting their user.

H2 - The Companion should migrate based on the context it perceives, the function it will perform (i.e. the particular embodiment required to perform the desired task), the role it should play in each embodiment (Kiesler and Goetz, 2002) and in response to a request from the user.

This allows the Companion to automatically select an embodiment best suited for the task without direct intervention from the user. The Companions presented in the video scenarios were implemented with this capability. This enabled participants to experience the scenarios and (through questionnaires) provide their feedback with regard to their approval, disapproval and suggestions for dealing with Companion's context of migration.

2.4 Companion Retention of Identity

If a Companion can be present in different devices, this might cause some confusion to the user regarding its identity (Tomlinson et al., 2006). An important aspect of Companion migration is the ability for a Companion to maintain its identity (Martin et al., 2005) and the user's belief that they are still interacting with the 'same Companion' in different embodiments (e.g., as it migrates from a humanoid robot to a zoomorphic robot platform). Most recent works addressing this issue use techniques such as displaying the owner's name associated with their Companion (Yasuyuki and Kenji, 2001) or representing the Companion with some consistent visual cues such as colour, markings, common features or class of objects (Martin, 2007). These techniques may be a good solution to assist users to easily identify their Companions in avatar form. However computing performance and resources available, and physical design restrictions can limit the behaviours, capabilities or features that might be characteristic to that Companion and therefore affect recognition by users (Martin et al. 2004).

This study addresses these issues with participants through two different migration scenarios. The first scenario involved Companions migrating between two similar virtual embodiments (avatars) with the same appearance. However, the Companions had different sounds and voices that indicated the migration process. The second scenario involved Companion migrating between two embodiments with different appearances (i.e. avatar and robot). However, the migration sound and voice was different, but a visual cue was used as an indication of who the Companion belongs to.

2.5 User's Privacy, Context of Use and the Functionality of Migrating Companions

Users' privacy is one of their main concerns, especially for Companions which have to learn about their users in order to serve them effectively. This is further aggravated when the Companion migrates to a different embodiment with private information. Therefore it is important to gain participants' perspectives on what kind of personal information they would allow their Companion to learn in order for the Companion to complete its task appropriately (i.e. trade off between user privacy and user requirements, and Companion functionality). Therefore the video scenarios were designed to elicit participants' responses, through an open-ended questionnaire, to the Companion seeking permission, functionality of the Companion migrating and personalisation of the Companion (Dautenhahn, 2004).

3 Study Design

The user study focused on Companion migration, taking account of the possibility that several Companions, associated to different users, may share the same set of available embodiments, or an embodiment that was different from what the user was used to. The study was conducted during the end of May 2009 at Wroclaw University of Technology in Poland in collaboration with the University of Hertfordshire in UK.

3.1 The Participants

A total of thirty-six participants, aged between 19 and 25 with a median age of twenty, took part in the study. The proportion of participants' genders was twenty seven males compared to nine females. Note, gender differences were not a key research question in our study, we thus did not attempt a gender balanced sample. Thirty of the participants were students of technical subjects while six participants studied non-technical subjects.

3.2 The Videos

Based on the above mentioned two hypotheses (*H1* and *H2*), two videos of relevant scenarios were produced; one focused on Companion migration between two avatars (virtual embodiments), the other focused on Companion migration between an avatar and a robot (physical embodiment). The Companions were able to search for people through a local area network (Scenario 1) and various rooms in the building (Scenario 2) by migrating to a physical robot if the person is not connected to the local area network.

3.2.1 Scenario 1 – Companion Migration between Two Virtual Embodiments

Companion migration between two virtual embodiments (Fig. 2) involved two students Lucas (i.e. Student 1 or S1) and Adam (i.e. Student 2 or S2) interacting

(a) (b)

Fig. 1. The two different embodiments used by the Companion in the videos. (a)The Companion's virtual embodiment (i.e. Xface) can be seen on the screen of the monitor, (b) the Companion's Physical Embodiment (i.e. Spunik) can be seen on the floor.

with their respective Companions, Alice (Companion of Student 1 or CS1) and Eva (Companion of Student CS2), which occupied avatars (Xface, see Fig 1.a) in their personal computers. S1 requested his Companion CS1 to deliver a message to S2. This involved CS1 asking S2 for appointments through the mediation of CS2 since CS2 was occupying the graphics character on S2's computer. If S2 agreed to interact with CS1, this involved CS2 informing S2 that it will relinquish the avatar for CS1 to migrate to. The migration process of both Companions was clearly indicated visually and aurally. The message was therefore delivered by CS1 to S2 in a way which took into account relations between S1 and S2.

After delivering the message both CS1 and CS2 migrated back to their previous embodiment. This scenario aimed to explore issues relevant to; i) decision of migration, ii) process of migration, iii) sharing the same embodiment by two Companions, and iv) retention of Companions' identities.

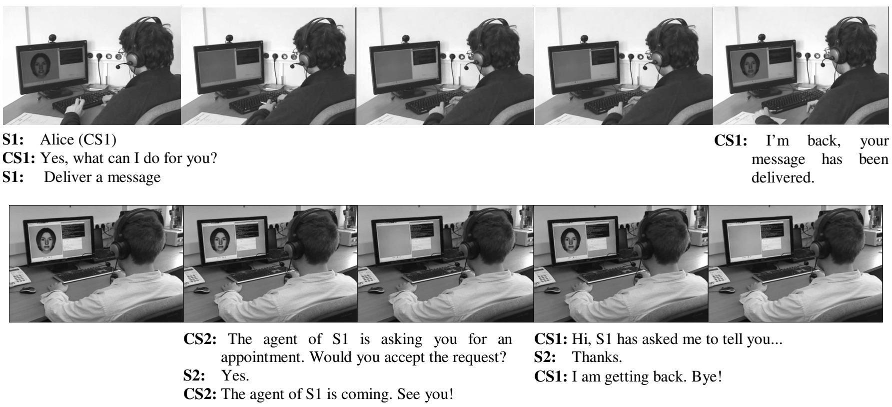


Fig. 2. Part of the narrative shown in the Scenario 1 video. The columns represent the event time-line, the text boxes show speech between the user and the Companions at that time-line event.

3.2.2 Scenario 2 – Companion Migration between a Virtual Embodiment and a Physical Embodiment

Migration between a graphical character and robot (see Fig.3) involved two students (S1 and S2) and a Teacher (T). S1 was interacting with his personal Companion CS1 which occupied Xface in his computer, while S2 was interacting with T in the robotic lab. S1 requested his Companion CS1 to deliver a message to S2. This involved CS1 migrating into a physical robot embodiment (Sputnik, see Fig 1.b) in order to search for S2 in the robotic lab. The migration process to and from both Xface and Sputnik was clearly indicated visually and aurally. Similarly, the message was delivered by CS1 to S2 taking account of relations with S1. This scenario explored issues relevant to; i) the decision of migration, ii) process of migration, iii) taking over different embodiments, iv) cooperation between the Companion and the robot's navigation system.

3.2.3 Companions' Behaviours and Embodiments Design Decisions

The process of migration with respect to the virtual embodiments was expressed by synchronised graphical animations and sound effects. The sound effects included sounds accompanying the departure and arrival of Companions in an embodiment. The sounds were based on Doppler shift effects, which provided a unique sound for each Companion. As with the avatar animation, the face of a graphical character disappeared from the screen just before a Companion vacated the avatar and only reappeared if a Companion migrated into the avatar. In the

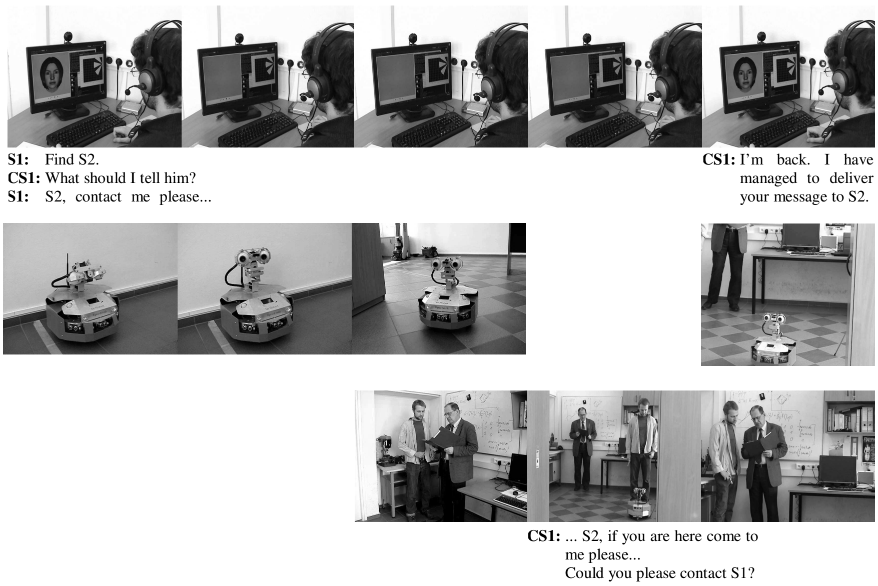


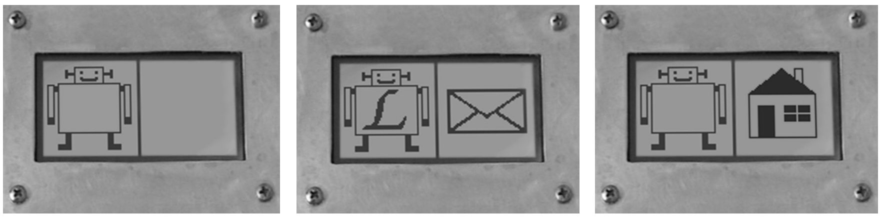
Fig. 3. Part of the narrative from the scenario 2 video. The columns represent the event time-line, the text boxes show speech between the user and the Companions at that timeline event.

trial the avatar was unique to a computer and not to any particular Companion. However the voice assigned to each Companion was unique.

The case of two Companions sharing a single avatar is very interesting from the perspective of migration. It allows the exploration of issues such as how participants felt about a Companion asking permission from its owner on behalf of another Companion and the ability of the participants to differentiate the two different Companions based on their behaviours and sounds (migration sound and speech). The latter issue is also relevant to users' perspectives of the Companion migrating into a robot (i.e. the robot had different speech synthesisers and quality of speakers).

In the case of the robot, the robot's physical behaviour (i.e. head lifted up from a bowed head posture), sound effects (unique to each Companion) and graphical cues were used for indicating the migration process. The graphical cues were expressed through a small display on the robot's LCD panel, which functioned as an electronic badge (see Fig. 4) and indicated the identity of the Companion, and also expressed its intentionality.

In the context of migration of a Companion to a robot, the identity and the intention of the Companion are very important. People should be aware of the robot's intentions and for whom the robot is performing the tasks. In this trial, the electronic badge was the best option for exploring these issues since the study was aimed at participants that were encountering these systems for the first time.



(a) (b) (c)

Fig. 4. Robot's LCD Information Panel. a) No Companion present and the robot has no task to perform. b) Lucas (S1) Companion is occupying the robot and has a message to deliver. c) Lucas's (S1) Companion has left the robot and the robot's basic function is activated to bring the robot back to its home location.

3.3 Experimental Procedure

The experimental procedure was conducted in the following stages:

Stage 1: Participants were introduced to the research, and questionnaires were handed out. They were asked to fill in the first page and not to go beyond the first page. They were also told that we are interested in their opinions and that there is no right or wrong answer to the questions.

Stage 2: The participants were shown Scenario 1 video; the Companion migrating between two virtual embodiments. After watching the Scenario 1 video, the participants were asked to fill in pages 2 and 3 of the questionnaire.

Stage 3: The participants were then shown the Scenario 2 video; the Companion migrating between a virtual and physical embodiment. Similarly, after watching the second video, participants were asked to complete the remainder of the questionnaire.

3.4 Questionnaire

The questionnaire consisted of 26 questions. They were divided into 6 pages where the questions on each page were formed to address a particular context. Page 1 of the questionnaire collected participants' demographic data. Page 2 and 3 dealt with Scenario 1, page 2 focused on participants feeling about the two Companions, while page 3 focused on issues of Companion migration between virtual embodiments. Page 4 and 5 dealt with Scenario 2; Page 4 focused on the Companion migrating into the physical embodiment (i.e. robot) while page 5 focused on participants' understanding of the robot's LCD information panel (which allowed the Companion to express ownership of the embodiment, and also showed the robot's and Companion's goal through the screen). Page 6 focused on the general issues of migration technology, such as possible usage other than that shown in the video and personalization of Companion.

Participants were told that they were under no obligation to complete the questionnaire, and to skip individual questions they did not wish to answer. However it would be appreciated if they completed the other questions.

4 Results

The results are divided into subsections addressing each research question proposed above (section 2). The results were based on descriptive analysis as the questionnaire was based on open-ended questions. Participants responses were categorised accordingly for analysis and their responses may not be mutually exclusive. Note that participants were asked to fill in the questionnaire voluntarily, hence it was expected that there were be some questions which certain participants may not have wished to answer.

4.1 Companion Migration between Computers

This sub-section addresses participants' feedback with regard to what participants saw in the Scenario 1 video.

4.1.1 Feeling about Their Companion Migrating Away

Overall, 25 participants felt positive towards the idea of their Companion migrating away from their computer, while 9 participants were negative. Twelve participants referred to how easy or difficult it was to accomplish tasks using the Companion in this manner, 10 participants referred to alternative methods of using the Companion to accomplish the task without migration. Eight participants addressed technical issues regarding migration and 6 participants referred to privacy issues in connection with Companion migration.

4.1.2 Feeling about Other's Companion Migrating into Their Computer

Overall, 24 participants felt positive towards the idea of another person's Companion migrating into their computer, while 9 participants were negative. The reasoning behind their responses was related to Privacy, Control and Ease of Use, with each having 16 references, 13 references and 9 references respectively.

4.1.3 Companion Asking Permission

Overall the answers of 26 participants suggested that they were satisfied with the way that the Companion asked for permission to migrate, while 8 participants were dissatisfied. The reasoning given by the participants were sorted into 5 categories: No added information, Politeness/Social Acceptability, Duration, Ease of Use, and Necessity of Permission.

Of the 26 satisfied participants, 8 of them indicated that Politeness/Social Acceptability was the main reason behind their satisfaction for the way the Companions asked for permission (i.e. the Companions were polite, acting quite natural and their behaviours were appropriate to the situation). Ease of Use was the main reason for 6 participants (i.e. used a simple question to get a straight answer, it was factual and understandable, etc.). Three participants highlighted the necessity of permission where they indicated their concerns for privacy (i.e. our own privacy is a very important issue; many questions are necessary, etc.). Nine participants from the No Additional Information category complemented the way permission was asked (i.e. it was good, it's ok, the way was decent, etc.)

Five participants who were dissatisfied with the way the Companions asked permission highlighted that the process took too long, while one participant proposed the use of a different modality (i.e. show an icon) to speed up the permission seeking process. Interestingly one participant actually preferred the Companion to migrate without asking for permission due to dissatisfaction with the way the Companion sought permission from the user.

When participants were asked to suggest a better way for the Companion to ask for permission, 6 participants suggested using different modalities (i.e. use a short text message if the user is busy on the computer, or otherwise use speech, but with a short status message like "question is being asked", highlighted by suitable characteristic sound, etc.). Two participants suggested that the duration should be shorter (i.e. shorter speech). Interestingly one of the participants even suggested this at the expense of not having the Companion seeking permission. Two participants suggested that the Companion should have a higher clarity of speech (i.e. basic vocabulary and speak slower) and one participant suggested the Companion should be more personal and have a better knowledge of its owner.

4.1.4 Differences between the Two Companions

Participants' feedback with regard to seeing the differences between the two Companions (Alice and Eva) suggested that the Companions' sound (i.e. migration sound, speech) and behaviours were the main characteristics that helped them differentiate the two Companions. Of the 16 participants who were able to differentiate the two Companions, 12 references were attributed to the Companions' sound (both migration sound and speech) and only 7 references were attributed to the Companions' behaviour. Sixteen participants were not able to distinguish between the two Companions, but only 4 participants provided their reasoning. All four references were directed at the similarity in the Companions' appearances, 2 references to the similarity of the Companions' behaviours and 1 reference to the similarity of the Companions' sound as contributing to their inability to spot the different between the two Companions. The rest of the 12 participants did not provide any reason as to why they did not spot the difference between the two Companions.

Twenty-nine participants suggested that a better way of helping them distinguish the two Companions might be by altering the Companion's appearance, sound, gender, behaviours and to allow some personalization options. Interestingly 22 references were made toward the Companions' appearances, 11 references to the sounds, 6 references for personality, 5 references for gender and 4 references to the Companion's behaviours. Six participants suggested personalization option where they can personalize their own Companion.

A majority of the participants (i.e. 11 participants) preferred Alice (CS1) over Eva (i.e. 3 participants) while 19 participants preferred neither or could not differentiate between the Companions. Interestingly, participants who have a preferences for one Companion over another, indicated that their preferences for a particular Companion were based on the Companions' sound (i.e. 10 participants) and behaviour (i.e. 8 participants). As expected none of the participants attributed their preferences to the Companions' appearances, since both Companions had the same appearance.

4.2 Companion Migration between Computer and Robot

This section addresses participants' feedback with regard to participants' perceptions and views on the Scenario 2 video.

4.2.1 Realisation of Companion Migration to Robot

Twenty-four participants were satisfied with the way the Companion migrated into the robot (as shown in the Scenario 2 video), while nine participants felt negative. Overall, twelve participants complemented the migration process and the idea of Companion to robot migration, 8 participants referred to the speed of the migration process, 3 participants referred to the robot's capabilities to provide extra functionality to the Companion, and 3 participants liked the Companion making use of simple technology (i.e. WI-FI for migration). Three participants suggested the idea of splitting the Companion into two entities i.e. "Companion should divide itself when migrating to the robot, so it would still be available on the computer for the user."

4.2.2 Retention of Companion Identity – Migration from Avatar to Robot

Interestingly, thirty-one participants could not see retention of Companion identity in the robot after the Companion migrated to the robot, while three participants could. A majority of the reasons given by the thirty-one participants were that the robot's voice was different from the Companion (13 references) and that there was no added information on the robot to indicate the Companion's identity (13 references). Of the three participants who noticed the retention of Companion's identity in the robot, two of them referred to the icon "*L*" on the robot's LCD panel, while one participant stated it was the robot's agency (i.e. the robot shared the goal of the Companion).

Although few participants acknowledged that the Companion retained its identity, it is interesting to note that the reasoning behind the most common reason raised for it not retaining its identity was the voice attribute. This may due to the fact that participants were used to the concept that the Companions have different sounds and maintained their individual sounds when migrating to different embodiments as shown in the Scenario 1 video.

4.2.3 Robot's Information LCD Panel

Participants' feedback with regard to the robot's LCD panel indicated that twenty-seven participants understood the message on the LCD panel while seven participants did not. Participants' descriptions indicated that eleven participants understood the meaning of the Companion related information on the LCD panel, while twenty-four participants did not. However, thirty-two participants understood the meaning on the Task related information LCD panel, while 3 participants did not.

In the last segment of the scenario 2 video after the message was delivered to S2, participants were asked to describe what they thought the robot and the Companion were doing. 14 participants indicated that the Companion migrated, then the robot's basic functionality drove it back to the base. Five participants indicated that the Companion drove the robot back to the base before migrating away. Three participants stated that there was no migration happening, but the Companion was communicating with the robot. Ten participants did not refer to the state of the robot and Companion (i.e. "they are waiting for a new mission", "they return to their initial position", etc.).

4.3 Participants' General View on Migration Technology

This section addresses participants' feedback with regard to the general use of migration technology.

4.3.1 Advantages and Disadvantage of Migration Technology

Speed, convenience, entertainment and added functionality were the major benefits thirty-one participants saw in the Companion migration between computers in the context of the Scenario 1 video. However 30 participants were concerned that such technology will create less human contact, annoyance, less privacy, be impractical and cause technical problems.

Based on 29 participants' feedback, the benefits of Companion to robot migration in the context of the Scenario 2 video were movement, locating people, physical tasks, security and convenience. The drawbacks of such technology in this context were low speed, cost, less human contact, annoyance, lack of privacy and security, and technical problems.

4.3.2 Perceived Applicability of the Migration Technology

When participants were asked if they would use migration technology in the future, nineteen of the participants stated that they would use the migration technology, while fifteen participants stated that they would not use the technology. The main reasons behind participants' decisions can be classified into seven categories:

Further development – Nine participants indicated that they would use the technology if it were improved (e.g. faster respond, clear speech synthesizer, reliable).

Direct Contact - One participant would use the technology to locate a person, while 7 participants indicated they would not use the technology because they prefer direct contact.

Usefulness - Four participants would use the technology because it would facilitate them in their daily life, while two participants would not use the technology because they could not see its usefulness.

Time – Three participants indicated that they would use the technology because it would be able to save them a lot of time (including multitasking).

Wide-spread Adoption – Two participants said they would use the technology when it became widely accepted.

Privacy – One participant said he would not use it because it would damage their private space.

No Additional information – Five participants did not provide any information regarding why they would use (3 participants) or would not use (2 participants) the technology.

Further exploration of the applicability issue for Companion migration between computers shows that 14 participants would use it for internet activities, four for large scale contacting and three participants to avoid personal contact. Fifteen other participants did not provide any information, or indicated they do not know how to use technology apart from the one that was shown in the Scenario 1 video. As for Companion-robot migration technology, six participants indicated that they would use it for housework, seven for shopping, three for object manipulation and five participants would use it for surveillance/observing. Similarly, 15 participants did not provide any information or indicated activities that are similar to what they saw in the Scenario 2 video.

4.4 Personalisation of Your Companion

Thirty-one participants said they would like to customize their Companion while 2 participants did not. Eighteen references were made toward customizing the Companion's appearance, seventeen toward the Companion's sound and twelve toward the Companion's (unique) behaviours or functionality. Seven participants would like to have some of their own characteristics being added to their Companion, three participants would like to model their Companion with a specific attribute (appearance or voice) of a well known character. Note that participants' customization preferences were not mutually exclusive.

Twenty-five participants said they would want their Companion to learn about them, while eleven participants did not. Participants indicated that they would like the Companion to learn about their general personal details (10 references), their friends and family (8 references), their scheduling (7 references), their health related issues (2 references), their hobbies (12 references e.g. music, movies), and their interaction preferences (6 references).

5 Discussions

One of the main aspects of Companion migration is the technology itself, where a personal Companion, which holds private information may migrate between different embodiments. This may not be an issue for the user if the Companion is migrating between embodiments which belong to the user. However there are situations where one's Companion may have to migrate to an embodiment belong to someone else, as illustrated in the first scenario. How would users of such technology feel about their Companion migrating to other's environment and what about another person's Companion migrating into their environment? The results from our study show that around 2/3 of our participants were positive towards the idea of their Companion migrating to other person's environment, and another person's Companion migrating into their environment. However it is interesting to note that participants' main concerns were more about the migration technology and its usability when asked about their Companion migrating to other's environment rather than privacy and control issues. These latter issues were their main concern when they responded to the question of another person's Companion migrating into their own environment. The situation of participants focusing on the technology and usability may be due to them seeing the technology for the first time, which has increased their curiosity about how the system works, rather than their personal concern toward such technology (Kanda et al., 2007).

One of the ways to tackle the privacy and control issue is by having the Companion asking for permission and approval. Results indicate that participants prefer the Companion to be polite and exhibit social acceptability when seeking permission. When asking permission, the questions should be short and precise, understandable in order to afford short responses. Various methodologies (i.e. icon, sound, short message) participants are already familiar with could be integrated to achieve the design principle of short, precise, understandable questions that require short responses from user.

The results have also suggested that apart from the Companions' appearances, the Companions' sound and behaviour can also be used to help participants identify different Companions. Interestingly for the Companion migrating between computer and robot (Scenario 2), the majority of participants were not able to see the retention of Companion identity in the robot because the appearance and sound of the robot was different from the avatar in which the Companion was residing previously.

These findings indicate that adaptation of the Companion's embodiment for a particular user should not be focused on the Companion's sound and appearance, as these attributes may change depending upon a particular embodiment. The Companion's embodiments should be designed to focus on users' attention to relatively constant attributes, such as behaviours which are developed through interaction histories and to consider sound and appearance attributes in conjunction with such behaviours. Furthermore, to communicate awareness and understanding of the previous interaction history between the user and the Companion, this can be done explicitly. The Companion should clearly identify itself verbally or symbolically to the user in each embodiment and also provide information specific to the Companion, in a manner analogous to that used by mediums claiming to be 'possessed' by specific individuals as described in (Anderson, 2003).

The robot's information LCD panel is a good addition to indicate which users' Companion is occupying a particular embodiment. This is particularly useful especially in an environment where Companions may share the same embodiment. It is also useful for the Companion to express its intentionality, especially for an embodiment that has no other means of expressing its intentionality (e.g. no voice or speech capabilities). These results seem to suggest that a task related information panel can be effective since the majority of participants seemed to understand the task related information LCD panel easily. This result may likely be influenced by the scenario context. Interestingly the Companion information LCD panel was not as effective as the Task information LCD panel. Only eleven participants understood the Companion related information LCD panel, where the character "*L*" in the robot icon represents that Lucas's Companion is occupying the robot. This may be due to participants not remembering the owner's name, and therefore not linking the "*L*" character to Lucas's Companion. This problem could be solved by labelling each LCD information panel appropriately to provide a context that will help the user to better understand the panel.

An important factor is to enable the owner of the Companion to personalise their own Companion's appearance, sound and behaviour in order for the Companion to be unique and easy recognizable, regardless of different embodiments restrictions that might mask some of the Companions unique attributes. As suggested by the participants, Companion appearances may be changed to look more like their owner, a famous character or even a combination of these. Customisable aspects of Companion sound could be tone, prosody characteristics, discourse marker or specific schematic sentences. Customisable Companion behaviours might include be movements and interaction patterns that are influenced by the Companion's personality or even evolve them through prolonged interaction with their owner. Participants indicated that they would want their Companion to learn about their hobbies, personal details, friends and family, scheduling, interaction preferences and health related issues.

Speed, technical problems, privacy and security, annoyance, and reduced human contact seem to be the main concerns with regard to migration technology.

A majority of participants have no objection to using migration technology if the technology is further improved (i.e. the Companion has a faster response when interacting with the user, has a clear speech synthesizer and is reliable). Their main concern is that such technology might indirectly prevent them having direct contact with another person. This is an important factor that was raised, but this may be due to the particular scenario presented (i.e. using the Companion to contact or search for another person).

6 Relevant Issues for Companion Migration

The results and discussion from this study highlighted many issues discussed above, in particular participants' main concern for the need of privacy and control of the Companion and its embodiments, and the desire for streamlining the process of migration.

In this section we propose a possible solution to these issues and link it to the Companion's decision for migration and selecting an embodiment. The whole process of Companion migration can broadly be divided into three processes: *Authorisation for Migration*, *Authorisation for accessing an Embodiment* and *Authorisation for accessing the data stored in the Embodiment*.

6.1 Authorisation for Migration

The decision for a Companion to migrate to a different embodiment is a complex issue and ideally should be based on the context and task the Companion is performing, and with explicit permission from the owner to authorise the migration.

However, the Companion should also be able to learn about its owner's decisions with regard to its migration authorisation through their interaction history and apply those decisions in future migration processes based on the context and task. For example, a Companion may have learned from previous experience that its owner always gives permission for it to migrate to the embodiment with a gripper when performing a fetch and carry task at home. Therefore, the Companion may automatically migrate to that embodiment without explicitly seeking permission to migrate when performing that particular task. Apart from learning through an interaction history, the Companion should also have the ability to use implicit permission from the owner. For example, if the Companion inhabiting an avatar (virtual embodiment) is asked by the owner to fetch a cup in the real world, it should understand that the request includes implicit permission for it to migrate to a physical embodiment to perform the requested task.

We therefore believe that the decision on selecting an embodiment, and authorisation for Companion migration can be divided into three categories. The first requires a Companion to seek explicit permission from the owner, at the second involves the Companion obtaining implicit permission through an owner's request, while the third involves the Companion obtaining permission through interaction histories.

In this study we observed two types of migration approval. In the case of S1, the approval for CS1 was by default and was a consequence of commissioning the task (i.e. implicit permission). In the case of S2 the conscious acceptance of the user was required (i.e. explicit permission) for CS1 to migrate into S2's computer.

6.2 Authorisation for Accessing an Embodiment

We suggest that there has to be a security protocol established for protecting embodiments from occupation by unauthorised Companions, similar to the security protocol established to stop unauthorised computers from accessing a Wireless LAN. We define three groups for the security protocol (LIREC Deliverable D8.1, 2009):

Long-term Authorisation – This protocol is usually applied to Companion embodiments owned by the host user. It allows a Companion to have full access to the embodiment over the long term if it has the Authorisation for Migration. This was demonstrated in Scenario 1 where CS1 do not need permission to migrate back and forth to the avatar on S1's computer.

Short-term Authorisation – This is to specially cater for a trusted visitor's Companion. The visiting Companion's owner may be a trusted friend or family member of the host user. Different layers of access are defined in this protocol with regard to categories of information stored in the embodiment which are available for access by a visiting Companion (e.g. the visitor's Companion may have access to information such as where cups are located in the house, but may not have access to the owner's private information). The duration of the access may be limited to a fixed period between a few days to a week or so. This was demonstrated in Scenario 1 where CS1 seeks permission from S2 to migrate into the avatar on S2's computer.

Task-based Authorisation – This applies to a visitor's Companion with whom the host user is not very familiar. The authorisation will be given only for the visiting Companion to perform a specific task, and it will only have access to information stored in the embodiment that is directly related to the task. Upon completion of its task, the Companion will have to migrate away from this embodiment. This was used in Scenario 2 where CS1 migrated into the robot to search for S2 in the robotic lab.

6.3 Authorisation for Accessing the Data Stored in the Embodiment

Due to the current technology, it is impractical for a Companion to migrate non-essential additional data with it during migration. Therefore embodiment specific information that the Companion may learn about the environment will be stored locally in the respective embodiments and disseminated to other embodiments during when the Companion and the embodiments are idle, and the user is resting (e.g. between 1am-4am). As a result, some data stored in the embodiments might be too sensitive for a third party's Companion (Syrdal et al., 2007). Therefore, it is essential to establish a security protocol to govern access to sensitive data. We propose a group based security protocol that is linked to the *Authorisation for Accessing Embodiment* security protocol. Therefore the data stored in each embodiment is categorised, based on sensitivity and task, before linking it to the proposed three authorisation levels for accessing the embodiment.

Acknowledgement

The work described in this paper was conducted within the EU Integrated Projects, LIREC (Living with Robots and interactive Companions). Funded by the European Commission under FP7-ICT under contract FP7-215554. We are grateful to university students from Wroclaw, who participated in the experiment. Special tanks to Jan Kędzierski and Adam Oleksy for assistance in the preparation and carrying out the experiment, Daniel Davies for back translation on the questionnaire, Krzysztof Tchoń and Roboert Muszyński for involvement in parts of the experiment.

References

Anderson, R.: Defining the Supernatural in Iceland. Anthropological Forum 13(2), 125–130 (2003)

Bickmore, T., Picard, R.: Establishing and maintaining long-term human-computer relationships. ACM Trans. Computer-Human Interaction 12(2), 293–327 (2005)

Chapman, D.W., Carter, J.F.: Translation procedures for the cross cultural use of measurement instruments. Educational Evaluation and Policy Analysis 1(3), 71–76 (1979)

Dautenhahn, K.: Robots We Like to Live With?! - A Developmental Perspective on a Personalized, Life-Long Robot Companion. In: Proc. IEEE RO-MAN 2004, 13th IEEE International Workshop on Robot and Human Interactive Communication, Kurashiki, Okayama Japan, pp. 17–22 (2004)

Goodrich, M.A., Schultz, A.C.: Human-Robot Interaction: A Survey. Foundations and Trends in Human-Computer Interaction 1(3), 203–275 (2007)

Imai, M., Ono, T., Etani, T.: Agent migration: communications between a human and robot. IEEE International Conference on Systems, Man, and Cybernetics 4, 1044–1048 (1999)

Kanda, T., Shiomi, M., Miyashita, Z., Ishiguro, H., Hagita, N.: A two-month Field Trial in an Elementary School for Long-term Human-robot Interaction. IEEE Transactions on Robotics 23(5), 962–971 (2007)

Kidd, C.D., Breazeal, C.: A Robotic Weight Loss Coach. In: Twenty-Second Conference on Artificial Intelligence, Vancouver, British Columbia, pp. 1985–1986 (2007)

Kiesler, S., Goetz, J.: Mental models of robotic assistants. Extended Abstracts. In: Conference on Human Factors in Computing Systems, pp. 576–577. ACM Press, New York (2002)

Koay, K.L., Syrdal, D.S., Walters, M.L., Dautenhahn, K.: Living with Robots: Investigating the Habituation Effect in Participants' Preferences during a Longitudinal Human-Robot Interaction Study. In: IEEE International Symposium on Robot and Human Interactive Communication, pp. 564–569 (2007)

Koay, K.L., Syrdal, D.S., Walters, M.L., Dautenhahn, K.: Five Weeks in the Robot House. In: Second International Conferences on Advances in Computer-Human Interactions, pp. 219–226 (2009a)

Koay, K.L., Syrdal, D.S., Walters, M.L., Dautenhahn, K.: A User Study on Visuallization of Agent Migration between Two Companion Robots. In: 13th International Conference on Human-Computer Interaction (2009b)

LIREC Deliverable D8.1, Foundations of migrating companions (2009)

MacDorman, K.F., Cowley, S.J.: Long-term Relationships as a Benchmark for Robot Personhood. In: The 15th IEEE International Symposium on Robot and Human Interactive Communication, pp. 378–383 (2006)

Martin, A., Duffy, B., O'Hare, G., Schön, B., Bradley, J.: Empowering Agents within Virtual Environments. In: Proceedings of the IEEE Systems, Man, and Cybernetics UK-RoI 3rd Workshop on Intelligent Cybernetic Systems. IEEE Computer Society, Los Alamitos (2004)

Martin, A., O'Hare, G.M.P., Duffy, B.R., Schön, B., Bradley, J.F.: Maintaining the Identity of Dynamically Embodied Agents. In: Panayiotopoulos, T., Gratch, J., Aylett, R.S., Ballin, D., Olivier, P., Rist, T. (eds.) IVA 2005. LNCS (LNAI), vol. 3661, pp. 454–465. Springer, Heidelberg (2005)

Martin, A.: Dynamically Embodied Virtual Agents, in School of Computer Science and Informatics. University College Dublin, Dublin (2007)

Newell, A.F., Carmichael, A., Morgan, M., Dickinson, A.: The use of theatre in requirements gathering and usability studies. Interacting with Computers 18(5), 996–1011 (2006)

Ogawa, K., Ono, T.: ITACO: Constructing an emotional relationship between human and robot. In: Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication, pp. 35–40 (2008)

O'Hare, G.M.P., Duffy, B.R., Bradley, J.F., Martin, A.N.: Agent Chameleons: Moving Minds from Robots to Digital Information Spaces. In: Proceedings of Autonomous Minirobots for Research and Edutainment, pp. 18–21 (2003)

Syrdal, D.S., Walters, M.L., Otero, N.R., Koay, K.L., Dautenhahn, K.: He Knows When You Are Sleeping. Privacy and The Personal Robot Companion, Technical Report from the AAAI-07 Workshop W06 on Human Implications of Human-Robot Interaction, pp. 28–33 (2007)

Tomlinson, B., Yau, M.L., Baumer, E.: Embodied mobile agents. In: Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems, pp. 969–976. ACM, New York (2006)

Walters, M.L., Syrdal, D.S., Dautenhahn, K., te Boekhorst, R., Koay, K.L.: Avoiding the Uncanny Valley – Robot Appearance, Personality and Consistency of Behavior in an Attention-Seeking Home Scenario for a Robot Companion. Journal of Autonomous Robots 24(2), 159–178 (2008)

Woods, S.N., Walters, M.L., Koay, K.L., Dautenhahn, K.: Methodological Issues in HRI: A Comparison of Live and Video-Based Methods in Robot to Human Approach Direction Trials. In: Proceedings of The 15th IEEE International Symposium on Robot and Human Interactive Communication, pp. 51–58. University of Hertfordshire, UK (2006)

Yasuyuki, S., Kenji, M.: AgentSalon: facilitating face-to-face knowledge exchange through conversations among personal agents. In: Proceedings of the Fifth International Conference on Autonomous Agents, pp. 393–400 (2001)

Security Robot Simulator

W.H. Hung[1], P. Liu[2], and S.C. Kang[3]

[1] Department of Civil Engineering, National Taiwan University, Taiwan, R.O.C.
[2] Department of Electrical Engineering, Tamkang University, Taiwan, R.O.C.
[3] Department of Civil Engineering, National Taiwan University, Taiwan, R.O.C.

Abstract. Building intelligent behaviors is an important aspect of developing a robot for use in security monitoring services. Simulating and testing the robot behavior in a virtual environment prior to producing the robot and then conducting practical experiments can greatly reduce the cost and duration of the testing process. This research proposes a framework for the simulation of security robots, called the *security robot simulator* (SRS), which is aimed at providing a fully inclusive simulation environment from fundamental physics behaviors to high-level robot scenarios for developers. Human simulation is also integrated into the robot simulator for simulating interactions between the security robot and human personnel. The simulator was implemented in Microsoft Robotics Developer Studio (MSRDS), a services oriented robotics platform composed of a simulation core and four decentralized modules: scenario event, patrol planner, robot unit, and civilian modules. The results show that the four modules fulfill the requirements of a security robot.

Keywords: security robot, robot simulator, MSRDS.

1 Requirements of the Security Robot Application

The technologies used in robotics have been widely researched and employed in various fields to assist people perform complex procedures in dangerous environments, such as in factories and hospitals. Originally robots were created to assist humans perform tasks which are highly repetitive. In recent years however, advanced robots are expected to be versatile and intelligent enough to handle different types of situations and interact with people in a wide range of different environments (Luo et al., 2007).

Humans today are also more concerned with the quality and comfort of their everyday lives. In addition, secure and safe living is a priority for most people. Hence, developments around security and service robots have become a growing research trend in the robot industry, especially as security robots will no doubt play an important role in our daily lives in the future (Su et al., 2004).

X. Wang (Ed.): Mixed Reality and Human-Robot Interaction, ISCA 47, pp. 111–132.
springerlink.com

The most common and basic service provided by a security robot is to patrol an assigned area without any external assistance. In more detail we define the requirements of an autonomous mobile security robot to be able to support the following basic requirements (as shown in Figure 1).

The requirements can be classified into four levels: control, function, task, and scenario:

- Control: This is the first and most fundamental part of the development; it controls and communicates with the robot's hardware, including the motors, sensors, computers, and electric power unit.
- Function: Functions consists of the basic "actions" which the robot must be able to perform. Functions include sensing, mapping, perceiving, localizing, motion planning (motions of moving platform or articulated arms), and path planning.
- Task: Tasks are assigned to the security robot by security guards or an artificially intelligent (AI) robot system. A task may consist of one or more functions together to accomplish a service. Tasks include exploring an unknown environment, tracking intruders, planning a patrol schedule, and execute a watchdog task.
- Scenario: Scenarios are groups of requirements that the robot utilizes to make decisions in particular situations. Actions based on scenario such as visitor guidance, injury alert and intruder detection, are categorized as human–robot interactions. Actions based on scenarios such as fire, blackout, and open entrance, which are relative to the robot's external environment, are categorized as environment–robot interactions. Requirements to deal with scenarios are high-level planning procedures and strategic decision-making algorithms that decide which task function should be performed.

An overall planning system based on all requirements developed in these four levels is needed for a complete and fully autonomous security robot. An intelligent robot system must also know how the robot should behave and when it must act autonomously, without artificial commands. This is the most important and difficult aspect of the system and requires successful integration of a number of the robot's requirements.

To research and test all these requirements on a physical robot would require budget, space, time, and troubleshooting knowledge of hardware. The equipment and sensors of a robot are often very expensive and can be easily damaged in an uncertain or complex environment, which is particularly true for an un-tested and un-verified control system. In addition, resource problems often exist for development of advanced robots, such as when multiple teams have to share a single robot, and integration can become even more complex (Microsoft 2009).

The advantages of using simulating a robot in a virtual environment prior to conducing real experiments are five-fold: 1) faster and more convenient setup; 2) better design exploration; 3) reduced cost of conducting experiments; 4) easier debugging; and 5) concurrent use. These advantages can expedite robot development and enhance the feasibility of the robot applications.

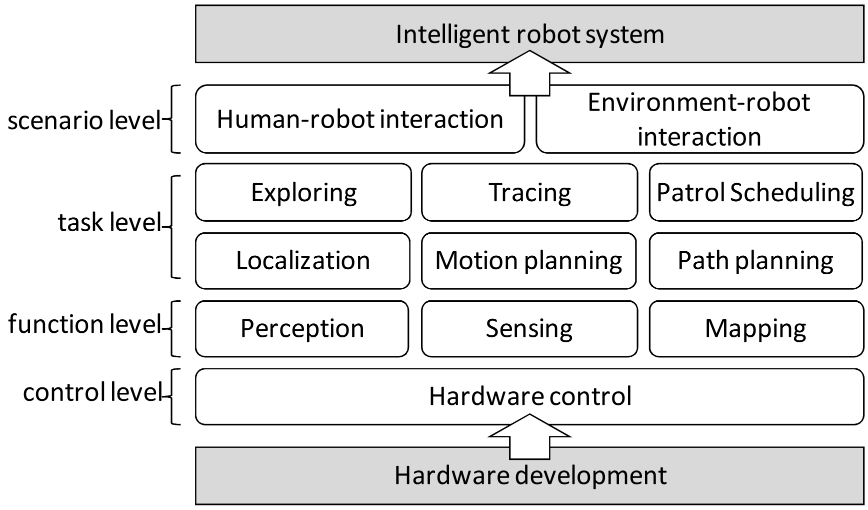


Fig. 1. Requirements for developing a fully autonomous and intelligent security robot

Different levels of the robot requirements matrix (as shown in Figure 1) require the simulator to support different levels of detail. However, previous research only covers a subset of these requirements. In addition, current algorithms used in simulation are neither expandable nor configurable for implementation on a real robot or other similar systems. Security robots are also different from other types of robot such as soccer-playing robots and cleaning robots in that one of their most important requirements is the capability to interact with humans. For example, during the day, a security robot would play the role of a guide for visitors within a designated area, which may be full of moving people; and at nighttime, the security robot will be required to patrol an assigned area to ensure there are no intruders trying to access the premises illegally. If an intruder were detected in the area, the robot would need to be able to respond to the situation by performing tasks such as tracking, blocking, or even attacking. To the best of our knowledge, there is currently no such simulator for security robot that combines the simulation of human behaviors such as pedestrian or intruder simulation. Hence, we propose a complete security robot simulator as a simulation environment that can support all the requirements presented in Figure 1.

This research proposes a system framework for security robot simulation, called *Security Robot Simulator* (SRS), which is expandable, and in which each of the components can work independently to increase the modularity of the developed program. We also include the simulation of human actions as a component in the simulator to provide users with a more complete testing environment for a security robot. The simulator also allows users to test and simulate their planning methods from the most fundamental physical motion of robots to more high-level scenarios in which robots may play.

2 Literature Review

Researchers have developed various types of security robots, each with different ad-hoc functions. Regarded as the world's first autonomous security robot, ROBART I was developed at the Naval Postgraduate School (Everett, 1982). Second and third generation counterparts, ROBART II and ROBART III, have also been developed (Everett and Gage, 1996). ROBART II employs 132 external sensors for navigation and intrusion detection; ROBART III is equipped with a Gatling gun style rotating barrel arrangement and is capable of autonomously navigating in semi-structured environments. Shimosasa et al. (1999) developed a guard robot which can perform visitor guidance and patrol functions. Birk and Kenn (2002) developed a mobile security robot that is visually controlled by humans via standard network technologies. A human-like security robot, MARVIN was designed to act as a security agent for indoor environments and was capable of interacting with people who have had little or no prior knowledge of robotic devices (Carnegie et al., 2004). Su et al. (2004) developed a low-cost security robot that was equipped with a multisensor-based system. Its second generation, WFSR2, was also built (Chien et al., 2005). Advanced security robots, such as Chung Cheng I (Luo et al., 2005) and NCCU Security Warrior (Luo et al., 2007), combine articulated arm actions and are able to support smoke detection and fire extinguishing capabilities. Many corporations have also developed their own service and security robots, and since 1982 some have been commercialized and employed at exhibitions and in malls (ALSOK, 2009; Dr. Robot, 2009; MobileRobots, 2009; Neobotix, 2009; ROBOWATCH, 2009; SANYO, 2009; SECOM, 2009; SKS, 2009; TMSUK,2009).

Earlier research in simulation predominantly focused on using mathematical methods and physics-based computational methods to simulate the dynamical motion and sensors of robots (Oh and Orin, 1986; Kuc and Siegel, 1987; McKerrow, 1989; Lemoine and Pape, 1991; Murphy et al., 1991). These simulators served to investigate and verify the control model and various robot systems, and cover the Control level of the four levels of requirements as shown in Figure 1.

More recently, researchers have focused on using computer graphic technologies for visualizing the simulation of robots to provide users with a higher-level testing environment (either task or scenario based) (Tarnoff et al., 1992; Kimoto and Yuta, 1995; Matsumoto et al., 1999). Rohrmeier (2000) used VRML to integrate interactive 3D graphics on the web to develop a web-based robot simulator. These simulation tools start to consider the Function and Task levels of our requirements. To build up a real-time interactive and realistic simulator, which allowed simulation of the physical motion of robots, researchers have begun to integrate real-time physics engines to increase the accuracy and reliability of simulations. Turnell et al. (2001) developed a simulation tool with a physical representation method called Simbot, a more reliable and advanced physics engine using constraint-based dynamics, ODE, which was integrated into robot simulators such as ÜberSim (Browning and Tryzelaar, 2003), Webots (Michel, 2004) and SimRobot (Laue et al., 2006). Meanwhile, a robot simulator, called USAR, developed using a game engine was proposed by Wang et al. (2003). A more extensible robot simulator with the PhysX physics engine, called VSE, was one of the major components of MSRDS (a development environment for robotics application) proposed by Microsoft (2007). Kucuk and Bingul (2009) developed an off-line robot simulation toolbox based on MATLAB.

Simulators, which utilize real-time physics and graphic engines, can cover not only the Control but also the Function and Task levels of the requirements. However these simulation tools aim to provide developers with a generic simulator for a number of purposes such as playing soccer, urban rescue, or home guarding. Users have to spend much time creating the Scenario level of the simulation environment. This research paper aims to provide a complete simulation environment which covers all the aforementioned requirements for security robots.

3 Concepts of SRS

The security robot simulator (SRS) is a framework for developing a simulation environment for security robots used in buildings and is a complete virtual testing environment suited for security and surveillance purposes. We will consider all aspects of the planning problem, including: basic robot body motion planning, path planning, perception, sensing, human–robot and environment–robot interactions. The SRS is composed of four independent modules plus a simulation core. The four modules are *scenario event*, *patrol planner*, *robot unit*, and *civilian*. The reasons we have chosen these four modules are as follows:

- Security problems are scenarios where the security robot has to be able to manage the situation. Typically there is neither a correct nor a standard solution to the problem. A simulator should therefore provide scenario events that may occur in the real world so that developers can have a high-level environment for conducting tests and investigations.
- Patrolling in an assigned area is the basic and original goal of developing a security robot. Security robots can work within a predefined area without a time limit and are equipped with advanced sensors to detect any dynamic changes inside the area. Therefore, a simulator should provide patrol quality assessment and reference path planning.
- Simulating detailed physical motion and behaviors of robots in a virtual environment is the original purpose of developing a simulation environment. Here, a robot unit module is used to provide developers with the means of building a physical model of their own robot and sensors in a virtual environment. Therefore, a testing environment for the function-level (physical motion and behaviors) and task-level (combination of functions) can be provided.
- Importantly, human factors influence the strategies and planning methods of a security robot, especially considering that humans can behave in unpredictable ways. Human–robot interaction (HRI) is a popular research topic, and simulation of humans in the patrol area is required.

The advantages of utilizing these modules in the SRS are as follows:

- Scenario-based simulation: The SRS can simulate security events such as fires, intruders, and injured humans.
- Self-workable and independent developable module: Each module of SRS can work, be developed, or be implemented independently.
- Consideration of human behaviors: This module provides the user with a complete simulation that considers human interactions while executing tasks.
- Extensibility: The isolated modules allow development to extend to other components and simulation environments.
- Reusability: Developed modules can be reused in other robot systems and simulation environments.
- Scalability: The SRS provides a framework for different levels of detail in simulations.

3.1 Simulation Core

The simulation core is the engine of the SRS and is responsible for the entire process, including both rendering and physics engines. The rendering engine, which is a basic requirement of a simulation system, is responsible for rendering 3D virtual scenes and objects to the screen. A rendering engine is typically developed based on OpenGL or DirectX. Existing rendering engines are capable of delivering realistic and content-rich visualization. Real-time visual effects also enable visual simulation of scenarios such as fire, smoke and blackouts.

Also key, the physics engine is responsible for the calculation of physical phenomenon, such as rigid body dynamics, collision detections, and contact responses. A real-time physics simulation engine has become a basic requirement in a virtual reality system. Various force interactions must be calculated to simulate robot motions and corresponding environment responses. Furthermore, various sensors attached to robots need to be simulated.

3.2 Scenario Event Module

The scenario event module is used to simulate the various scenarios encountered while the security robot is in service. Security robots can now support a wide variety of functions and are not limited to patrolling. Some advanced security robots claim that they have the ability to detect and extinguish fire, guide visitors to their destination, as well as deter intruders. Scenarios such as fire, however, are difficult to set up in order to test the robot's fire extinguishing capabilities; therefore, the scenario event module aims to deliver a plausible visual simulation of these scenarios to provide a safe and cheap testing environment. For example, the module can simulate fire, capture the robot's sensing ability with minimal or no light (to simulate a blackout), and simulate any other scenarios where a human is involved, such as detecting an injured person or an intruder.

3.3 Patrol Planner Module

The patrol planner is the basic component in a security robot simulator because patrolling is the security robot's core function. The patrol planner is a high-level strategic decision maker, which directs security robots with a pathway to traverse in a given environment.

In practice the most common method of defining the patrol problem is using checkpoints with time-windows. In other words, the human supervisors first must define several checkpoints in an assigned patrol space (an area to be patrolled) and a time-window (a range of time) to each checkpoint. Then the supervisor plans the patrol path according to the defined checkpoints to accomplish the patrol task; that is, each checkpoint should be patrolled at least once during its allocated time-window.

Advanced patrol-planning uses computational technologies, whereby the area is represented as a graph. This is in contrast to the typical continuous area patrolling. The nodes (checkpoints) correspond to specific locations and the edges of the possible path (Almeida et al., 2004). Instantaneous node idleness is a measurement index of patrol path and is defined thus: for a node *n*, cycle *t* is the number of cycles, which have elapsed since its last visit before *t*, and can be alternatively thought of as the number of cycles that node *n* has remained unvisited.

To generate a patrol path automatically, a fast patrol path planning method must be developed. Calvo and Cordone (2003) described a heuristic approach for the overnight security service problem by redefining the patrol problem as multiple travelling salesman problems with time windows (MTSPTW). Chevaleyre (2004) then developed cyclic strategies and partition-based strategies using the TSP solver to deal with single and multiple agents patrolling a predefined area.

In this research, we classify patrol planners into two basic types: *passive patrol planners* and *active patrol planners*. A passive patrol planner can inform the decision maker regarding the patrol status of the robot and assist in planning patrol strategies. This is aimed at defining the problem and status of the patrol and may include characteristics such as completed checkpoints, late patrols, localization of patrolmen or an emergency at a checkpoint. By comparison, an active patrol planner utilizes the information of the patrol status to find solutions for a patrol path automatically. An active patrol planner can improve the efficiency and mobility of the patrol strategies by dynamically re-planning as the situation changes. Since there are various types of autonomous robots developed for security purposes, the demand of high-level planning such as patrol path planning is gaining significance in the development of security robots.

To simulate the patrol scenario in a security robot simulator, both a passive patrol planner and an active patrol planner should be included, to allow users to test and verify their patrol planning algorithms.

3.4 Robot Unit Module

The robot unit module is used to test and verify motion control and high-level planning algorithms. In addition, the physical behaviors (including mobile platform, articulated arms, sensors, and collision geometry representations) are simulated so that users can build motion-planning methods and monitor sensors in a virtual environment prior to any physical experimentation.

The robot simulator includes two parts: the physics model definition and the robot status visualization. To simulate robot hardware with a physics engine, the physics model of the robot has to be defined prior to construction of the physics engine. During the simulation process, the information of sensors attached to the robots and other control data such as velocity, power of motors, or rotations of each joint need to be considered. Using this module, developers are able to modify parameters easily and debug accordingly.

3.5 Civilian Module

The civilian module is used to simulate a robot interacting with the motions and behaviors of humans by using information it extracts from the surrounding environment. Since security breaches are usually caused by humans, a security robot's ability to interact with humans and the environment is crucial.

Human–robot interaction (HRI) has recently become a trend in robot development. General HRI research includes human detection and response, and motion planning. Service problems under HRI in buildings are intruder detection, visitor guidance, and normal user behavior analysis. Steinfeld (et al., 2006) discussed the most basic functionality of robots, which is their ability to avoid objects that are moving, such as humans.

Further to the human factors, modern buildings have various sensors installed, such as cameras, smoke sensors, and door and window sensors. The dynamic data retrieved from these environment sensors are known as *building information*. By analyzing and utilizing the building information, we can efficiently manage and define a security robot's patrol strategies.

To satisfy the scenario-level requirements, the human and environment sensor simulations must be included in the civilian module in order to be able to provide a testing environment for human and building information interaction. Robots currently rely on data retrieved from cameras and distance sensors (such laser range finders and sonar sensor) to detect human behaviors and features such as movements and appearance (e.g., body shape and face pattern). Therefore, human simulation should include movement, physical shape, and visualization of bodies and faces.

Pedestrian simulation (or crowd simulation) is the simulation of individual or crowd movements. Interestingly, this has been researched and discussed over the past decade for city design and emergency evacuation. These researches use social forces modeling, agent-based reaction, cellular automation, or the potential field method to calculate the walking path of humans in a field automatically (Pelechano, 2007). Here, we combine the aforementioned technologies to simulate the movements of humans for a security robot simulator. In addition, physical shapes of humans are detected by the robot, whereas information of distance sensors predict the movements of objects and execute any avoidance actions; visualization of human bodies and faces are captured by the robot and the building's virtual cameras, allowing developers to build and test their recognition and detection algorithms.

3.6 Cooperative Model between Components

Figure 2 shows how the four components are isolated from each other, making them modular and therefore easily adapted to different problems and scenarios. Each module can work independently (using its own interface) or connect directly with the simulation core to simulate in a 3D virtual environment. A complete simulation using all four modules can be implemented with the simulation core, where the patrol planner and civilian modules continuously feed back and update information such as checkpoint status and the pedestrians' positions. The scenario events module provides the necessary resources to allow the simulation core to visualize the scenario. The robot unit module can then place the defined virtual security robot in the simulation core to simulate and test its algorithm and planning methods using the scenario provided by the other modules.

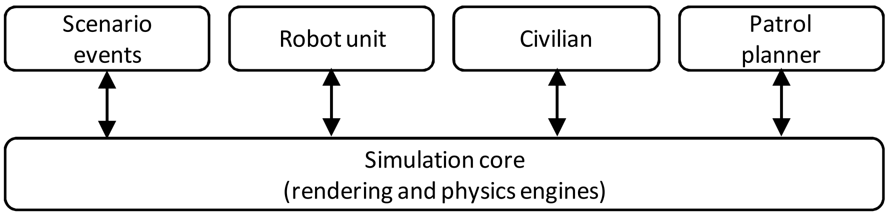


Fig. 2. Architecture of the SRS

4 Implementation

To implement the proposed simulator, we used Microsoft® Robotics Developer Studio 2008 (MSRDS), a Windows-based environment used by both academic and commercial developers for creating robotics applications for a variety of hardware platforms (Microsoft 2009).

4.1 Simulation Core Using MSRDS

The MSRDS platform provides a runtime and software structure for developing robot applications. The MSRDS development team incorporated various new programming concepts and functionalities into the software design, allowing the MSRDS to not only control the robot but also any other service-based equipment. Furthermore, MSRDS is tightly coupled with the Microsoft Windows® operating system, and because Microsoft Windows supports a wide range of software programs, MSRDS can easily be incorporated into the code base of Windows, making development easier and more flexible. Developing a robotics system based on this structure decentralizes the system and makes the components extensible and reusable. In other words, one of the main focuses of the development process is modularity. The user can develop and test the robot using the visual simulation environment (VSE) in a virtual environment.

4.1.1 Runtime and Structure of System

Two major features of MSRDS are decentralized software services (DSS) and concurrency and coordination runtime (CCR).

The DSS is the architecture for software development. It provides a state-oriented service model for building high-performance and scalable applications. *State* refers to each parameter of the robot service component, and is unique within the MSRDS platform. When any service component changes state, other service components will be automatically triggered to observe the latest state information and users need not worry that the system information is not up-to-date. The DSS service components can be also packaged and ported to another robot's control program for re-use, taking advantage of the highly object-oriented programming design concept. DSS service components also include a network component. MSRDS programmers can also use XML to access a robot's state information and publish it online for sharing with other programmers.

On the other hand, the CCR addresses the need of service-oriented applications to manage asynchronous operations, deal with concurrency, exploit parallel hardware and overcome partial failures. The concurrent programming technique is not uncommon, with one example being the multi-threaded processing technique used in operating systems of personal computers. However, one difficulty with using concurrent programming techniques is that a very high level of technical ability is required for actual implementation, which can be elusive even for people with a computer science or software engineering background. Fortunately, CCR enables programmers to write concurrent programs easily and intuitively and also makes debugging and data management more convenient (Microsoft, 2009).

Due to these two features, a simulator developed based on MSRDS is loosely coupled and has the potential to directly apply the developed algorithm and planning method to a real robot by using an extended service.

4.1.2 Visual Simulation Environment

Microsoft utilizes its rich experience in the gaming industry from the game development platform XNA to develop the Visual Simulation Environment (VSE). The VSE integrates XNA and PhysX, the rendering engine and physics engine respectively.

The XNA is a framework based on DirectX and is generally used for developing computer games and Xbox360 games (Microsoft, 2008). XNA has several technical advantages: it was originally designed for gaming purposes and provides excellent functions for user inputs and Internet connections. The development environment of XNA supports various formats of 3D models, 2D textures, and audio effects, minimizing issues where the programmer is restricted by problems associated with file transfers, which expedite the development process. Furthermore, virtual objects in XNA are all rendered using a graphics processing unit (GPU) shader, a small program that executes inside the GPU (Lobao et al. 2008).

Meanwhile, PhysX is a physics engine originally developed by a company called AGEIA, which was acquired by Nvidia in 2008. PhysX applies position based dynamics (PBD) to simulate physical behaviors in real-time to provide not only rigid body simulation, but also soft body, fluid, and cloth simulations (Nvidia, 2009). The PhysX engine has potential use in real-time simulation due to its efficient solver and high stability derived from PBD methods. The controllability of PBD allows the simulation of dynamic joint attachment and detachment without error propagation. With this advantage, contiguous actions can be simulated.

The VSE supports various well-known virtual sensors which are developed using XNA and PhysX, including laser range finders, cameras, bumpers, sonar, light (brightness) sensors, color sensors, compass sensors, infrared range, and GPS. The VSE provides a universal development environment for simulations that allows users to create and build up their own simulator and virtual hardware such as robots and sensors.

4.1.3 Graphics User Interface Using WPF

Besides the 3D simulation environment, MSRDS also combines the windows presentation foundation (WPF) for developing user interfaces. WPF is a graphical subsystem for rendering user interfaces in Windows-based applications. It provides not only the user interface but also 2D and 3D drawing, fixed and on-screen documents, advanced typography, images, animation, data binding, audio and video (Microsoft, 2009a). All the user interfaces of each module in this research were built using WPF.

4.2 Scenario Events Module

The scenario events module visualizes the scenario using various visual effects. We implemented three scenarios for the scenario events module; they are *fire accident*, *patrol scenario*, and *injury of people*. Fire accident is implemented using particle system and shader technologies that are generally used in computer graphics to represent an object on fire over a given time period, as shown in Figure 3.

Patrol scenario is used to visualize the degree of safety in a certain region as it changes over time, as shown in Figure 4. The green color shows that the region has a most safe state; otherwise the red color indicates a region in the most unsafe state, and which needs to be patrolled immediately. Users can define and visualize their own strategy for maintaining safety by using the functions provided, or alternatively, the definition provided by the passive patrol planner of the patrol planner module.

Injury of people is implemented by an animation of a person falling, as shown in Figure 5. Injury detection is one of the most important services provided by an advanced security robot, and can be trained by having people wear sensors and accelerometers to detect and analyze posture and status.

These scenarios can be edited and inserted into the VSE by users. When the scenarios occur, the module will automatically send messages to the VSE. Other services or modules can subscribe to these messages and respond or react to the scenario accordingly.

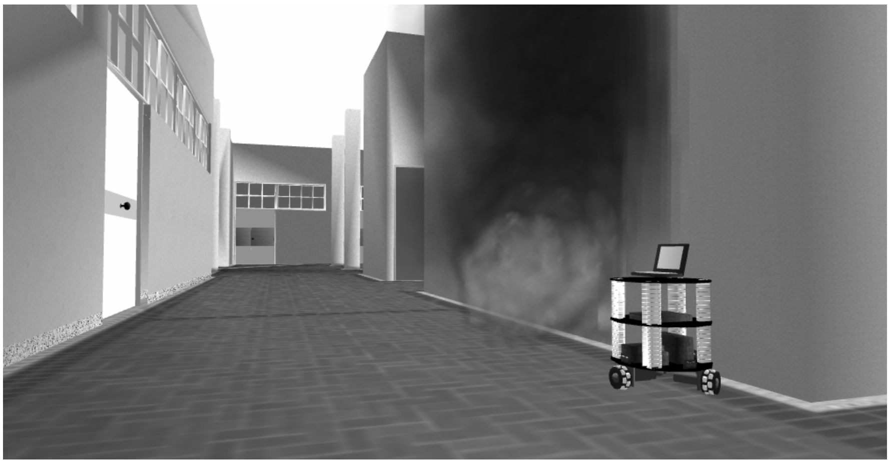


Fig. 3. A screenshot of fire simulation in the scenario events module. The robot which is proposed by Gu (2008) is built into the virtual environment.

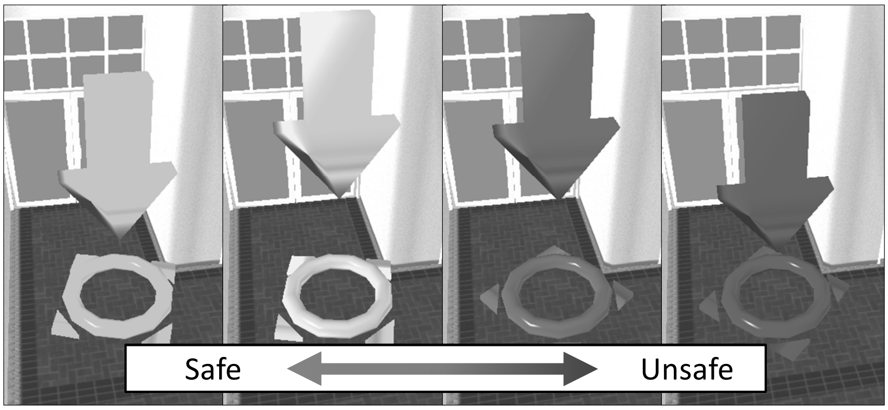


Fig. 4. Degree of safety of a certain region visualized by an arrow and ring with different colors in the scenario events module

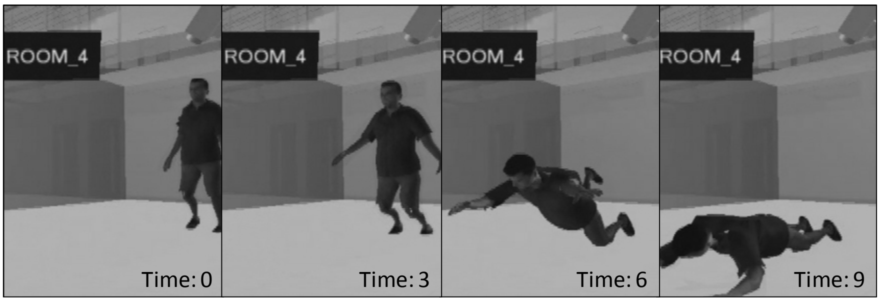


Fig. 5. Scenario of an injured person falling down as animated and visualized by the scenario events module

4.3 Patrol Planner Module

We developed a patrol planner module called *patrol path planner* (PPP). The PPP includes a passive planner and an active planner, and provides developers with a patrol scenario, quality measurement, and reference patrol paths.

4.3.1 The Passive Patrol Planner

We implemented a passive patrol planner based on the work of Hung (2008). In that experiment, a patrol area is regarded as a graph. A node is called a patrol region (a checkpoint) and is defined by the user. The patrol region is classified into three types: exit/entrance, property, and common region. Each region has an allowed-vacant-time (AVT) parameter defining the amount of time the region can be vacant. Patrol priority (PP) is the priority of a patrol region to be patrolled at a certain time and is calculated according to the type of the region and its remained-vacant-time (RVT), which is the remaining time of the region for patrolling. The followings are the equations for calculating these three types of patrol regions:

$$P^{ee}(r) = (1.1)^{-r} + 1 \tag{1}$$

$$P^{property}(r) = (1.2)^{-r} + 1 \tag{2}$$

$$P^{common}(r) = \{ \begin{matrix} (1.1)^{-r}, r \geq 0 \\ (0.1) \cdot r + 1, r < 0 \end{matrix} \tag{3}$$

where r is the RVT. This method provides an index of the robot's patrol path quality, denoted as *average patrol omission* (APO). Patrol omission (PO) is the value of PP when RVT of a patrol region is negative, representing the failure to patrol the region. Hence, decision makers then are going to minimize the following objective function:

$$f(t)=\sum_{t=0}^{n}\sum_{i=0}^{n}o_i\left(r_{it}\right) \quad (4)$$

where o_i is the function of the PO value associated with its r_{it} (RVT) of the i-th patrol region at time t. A lower APO value means that the planned patrol path can perform a better patrolling quality.

4.3.2 The Active Patrol Planner

Based on the passive patrol planner definition, we implemented an active patrol planner using the simulated annealing (SA) method to provide a reference patrol path. SA is a probabilistic meta-heuristic algorithm that accepts search movements, which temporarily produces degradations in a current solution found to a problem as a way to escape from local minima (Oliveira and Vasconcelos, 2008). Because the time of the objective function has no limitations, it is difficult to find a solution under such a condition. Therefore, we simplify the problem and try to find a path that can minimize the APO and cover all the patrol regions such that each region will be patrolled at least once. We begin by randomly assigning a path as s, and then proceed to find a new solution s_a which is also the neighbor of the s. If the $f(s_a)$ is smaller than $f(s)$, which means the new path is better suited as a planned patrol path, then we will accept it and replace the original s. Otherwise, we will accept the new path depending on the probability computed by the Metropolis Criteria, which is shown in Eq. 5, where T is a temperature parameter. This mechanism means that the system has a higher change of accepting a worse solution in order to avoid the local minima. The system will keep executing this process until a predetermined number of iterations have been reached (Oliveira and Vasconcelos, 2008). The method can also provide a high performance (low-APO) patrol path for reference.

$$p_{accept}=e^{\left.f(s)-f(s_a)\right/T} \quad (5)$$

Multiple robots patrolling together is another trend that has appeared in recent patrolling problems, as it can increase the coverage and robustness or the patrol (Hazon and Kaminka, 2005). We therefore have also developed a heuristic method for patrol path planning using multiple robots, which we have called *path evolution* in the active patrol planner. The concept of path evolution is that the robots repeatedly and sequentially plan their patrol paths depending on other robots' paths, and to keep improving and evolving their previously planned paths.

For example consider that we have three robots (r_1, r_2 and r_3) patrolling together in an area. In the first generation of the evolution, the robot r_1 plans its path via the SA method; the robot r_2 then plans its path according to the status of PP when r_1 is patrolling; finally, r_3 plans its path when r_1 and r_2 are patrolling. In the next generation, r_1 uses its current planned path as the initial solution in SA algorithm to evolve itself by re-planning its path according to the status of PP when r_2 and r_3 are patrolling. The patrol path of each robot will then keep being improved as it continuously executes this evolution process.

Path evolution is fast and effective method that can generate a quality reference path for either single or multiple robots.

4.3.3 Function Provided by PPP

The PPP provides a user interface for creating and editing the patrol area using the definition of a passive patrol planner, as shown in Figure 6. The patrol path planned by users can be executed and evaluated using the stand-alone PPP. Patrol regions to be patrolled by a virtual security robot can also be inserted into the VSE for 3D visualization, as shown in Figure 4. When the virtual security robot patrols a patrol region, it will update the information to the PPP and reset the RVT of that region. The reference path provided by the active patrol planner can help the user modify and improve patrol strategies.

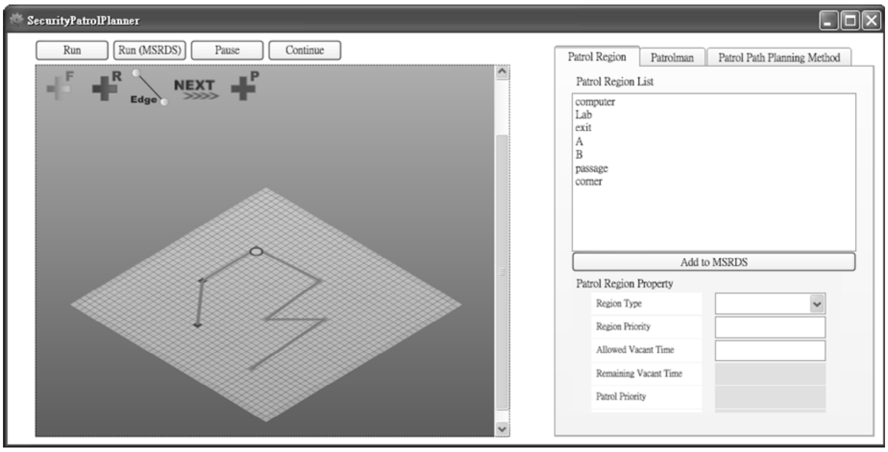


Fig. 6. The user interface of the PPP

4.4 Robot Unit Module

We implemented the intelligent security robot proposed by the Economic Bureau project "The Development of Integrated Intelligent Robotics System" in Taiwan (NSC 95-EC-17-A-04-S1-054) into the robot unit module. This robot consists of a mobile platform (two normal wheels and a caster wheel) and two articulated arms (each with seven degrees of freedom), as shown in Figure 7a. The robot is equipped with a laser range finder, a camera, a smoke sensor, and an array of microphones. We modeled the robot using multi-body dynamics shown in Figure 7b and 7c (Hung and Kang, 2009). A user interface was developed using WPF to visualize the data of the robots and to provide a control panel of the developed robot. The user interface currently supports data visualizations of cameras, laser range finders, motor speed, and joint angle of articulated arms. The user can insert the robot into the VSE to simulate and control the virtual robot from the robot unit module instead of using the VSE interface.

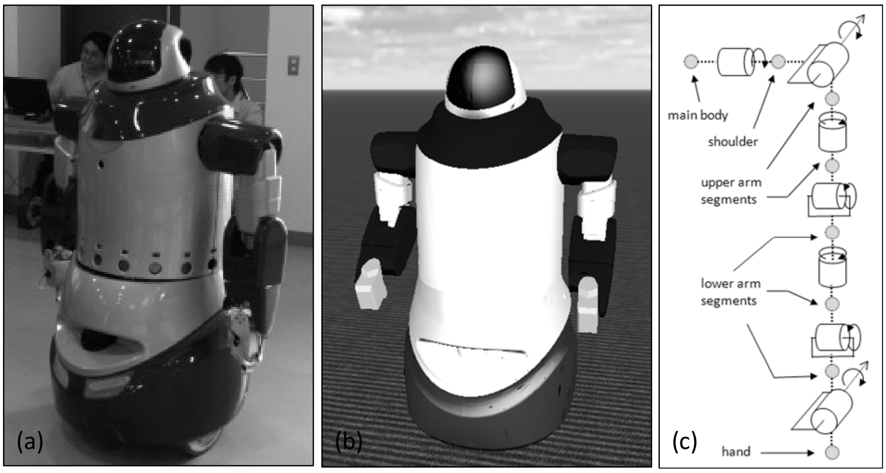


Fig. 7. The intelligent security robot: (a) actual robot; (b) screenshot of virtual security robot in VSE; (c) the articulated arm modeled and represented using multi-body dynamics

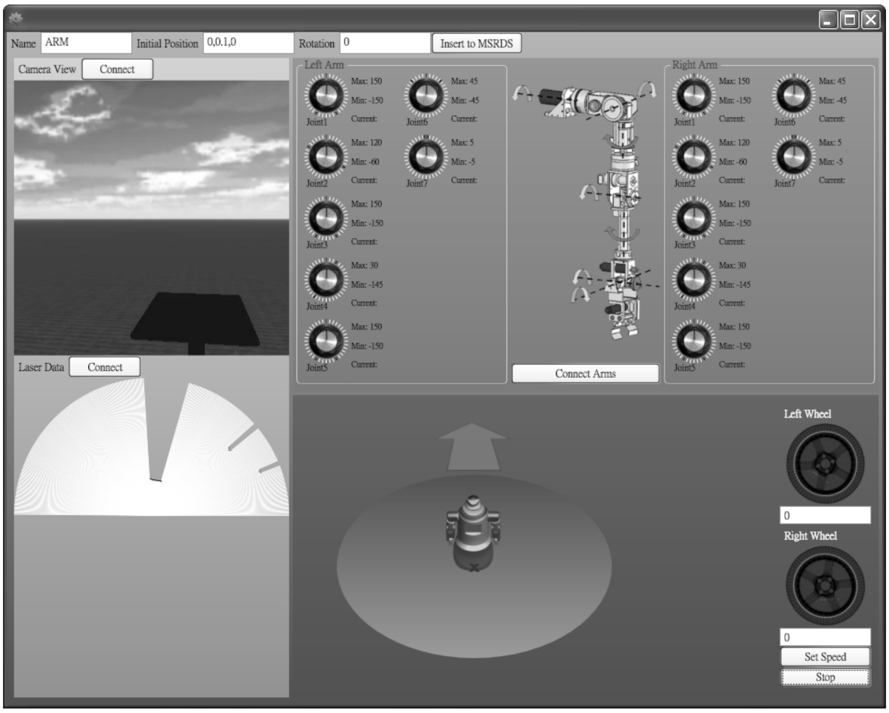


Fig. 8. User interface of the developed robot unit module

4.5 Civilian Module

The civilian module includes a pedestrian simulator, a physics model, and a visual model. A pedestrian simulator calculates the position of pedestrians at each time step and updates the information of the physics and visual models in the simulation core. The physics and visual models of humans are constructed for detection by the robot's sensor.

4.5.1 Construct Visual and Physics Model of Human

A visual model of a pedestrian should include a realistic appearance and a walking animation. The 3D model and animation of a pedestrian can be easily built using existing software packages such as 3DS Max, Maya, or Blender (Autodesk, 2009; Blender, 2009). A physics model of a pedestrian is a virtual doll composed of geometric shapes used by a physics engine for collision detection. We constructed the physics model of a human by using a simplified box to represent the body and sphere to represent the head. Figure 9 shows s screenshot of pedestrians being detected and visualized by a laser range finder in VSE.

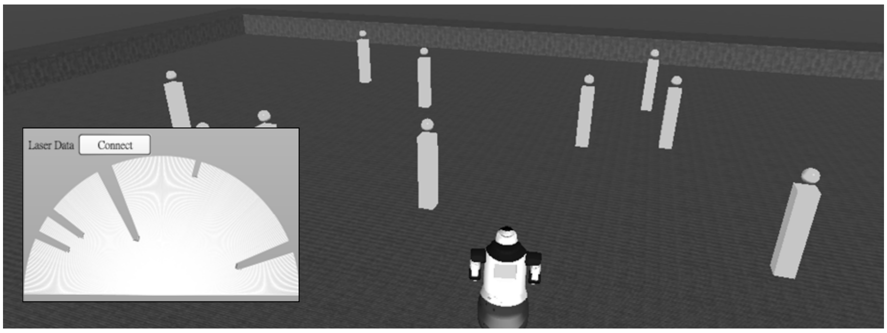


Fig. 9. A screenshot of a pedestrian's physics model and a laser range finder visualization in VSE

4.5.2 Method for Simulating Pedestrians

We developed an agent-based potential field method to simulate the pedestrian's movements. Traditionally, the potential field method has been used as it has the advantage of pre-calculating a map and automatically simulating the flow of crowd movement. However, this method omits individual pedestrian behaviors and any mutual interactions. Combining the potential field method with the agent-based method removes this disadvantage and the two combined are able to simulate pedestrians as well as the interaction between them.

The movement of a pedestrian in an environment is driven by three forces: a field force, a correlation force, and an individual force, as shown in Figure 10a. The field force is the force that attracts the pedestrian to a specific target. This force is pre-calculated and is consistent throughout the simulation. An area may have one or more field force maps according to the destination of pedestrians. To simulate differences between pedestrians, which are using the same force map, we multiply the influence by the field force with a field coefficient. The correlation force of a pedestrian is the force exerted by another pedestrian when that pedestrian crosses into the visual field of the original pedestrian. Figure 10(b) presents a correlation model that has three parameters: visual angle, effective distance, and safe distance. When someone steps into the safe area of a pedestrian, the pedestrian will stop until the safe area is clear. When two pedestrians are correlated, they start to interact, and a correlation force begins to act on one another; the individual force is the force that sums up the three forces of a pedestrian in the previous time step. The initial value of this force is zero. When the pedestrian starts to move, then all three forces are summed together at each time step.

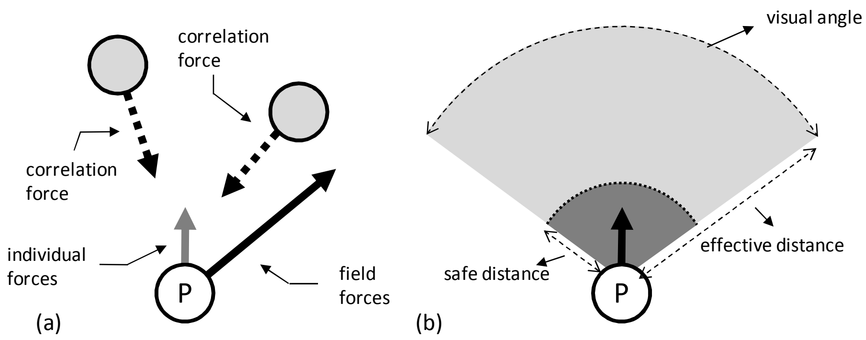


Fig. 10. The pedestrian moving strategy: (a) three forces that trigger the pedestrian; (b) correlation model

In this research, we use harmonic functions to generate the potential field. This method avoids the local minima, which would otherwise exist if calculated by Arkin's potential field (a widely used method for generating potential field by Dapper et al., 2006; Faria et al., 2006). This method can also deliver a real-time pedestrian simulation of up to two thousand simultaneously moving in an area; and, the user can dynamically add targets and pedestrians during the simulation.

4.5.3 Functions Provided by Pedestrian Simulator

The pedestrian simulator (PS) includes a pedestrian panel and a physical and visual model of a pedestrian for the simulation core. A screen shot of the pedestrian panel is shown in Figure 11a. Users can edit the map showing where pedestrians walk and assign either one or more exits to the area (destination targets of the pedestrians). The pedestrians in this area will tend toward these exits. Users can also define the number of pedestrians, and their individual parameters such as position or velocity. The panel was also built using WPF as a service. The pedestrian simulator can work independently by visualizing the simulation in 2D animations and by inserting the pedestrian model into the

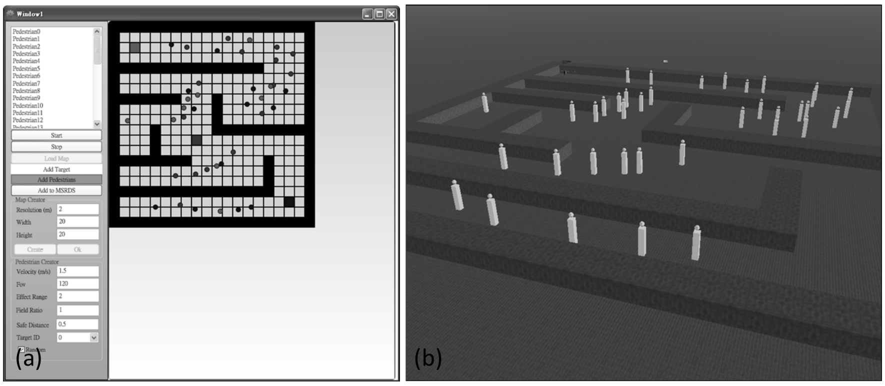


Fig. 11. Screenshot of the pedestrian simulator: (a) pedestrian editor; (b) pedestrian simulation in VSE

simulation core (which updates the position of each pedestrian to provide a testing environment with moving humans for the security robot). Figure 11b shows a screenshot of pedestrian simulation in VSE.

4.6 Discussion

These modules are services developed using CCR and DSS. Communication between modules and the simulation core is straightforward and is easily carried out by sending messages and data. Each module is isolated and can be extended and reused for other purposes and by other systems. The proposed modules provide the user with a complete testing environment for security robots taking into account factors such as human simulation. In this work, scenario events are simulated by using only visual effects and message notification, more detailed simulation such as fire, temperature and sound can be developed based on this platform by future research; the civilian module currently includes the simulation of pedestrians, which is used to simulate pedestrian movements for application testing such as path prediction.

5 Conclusions

This research presented a framework for simulating a security robot, called Security Robot Simulator (SRS). It includes four basic modules: scenario events, patrol planner, robot unit, and civilian. The modules can be run together by a simulation core to provide the simulation of robot applications in a realistic virtual environment. Since SRS integrates both human simulation and robot simulation, it is ideal to test the interaction between human and robots. SRS was implemented based on the simulator of Microsoft Robotics Development Studio (MSRDS). By utilizing the service-oriented architecture of MSRDS, this simulator can be easily extended to simulate and visualize other behaviors of robots. The developed modules can also be integrated with the actual robot for other advanced applications and purposes.

Acknowledgement

The authors would like to thank the Technology Development Program for Academia, Ministry of Economic Affairs, Taiwan, for their financial support under the project "The Development of Integrated Intelligent Robotics System" (contract no: 95-EC-17-A-04-S1-054).

References

Almeida, A., Ramalho, G., Santana, H., Tedesco, P., Menezes, T., Corruble, V., et al.: Recent Advances on Multi-agent Patrolling. In: Bazzan, A.L.C., Labidi, S. (eds.) SBIA 2004. LNCS (LNAI), vol. 3171, pp. 474–483. Springer, Heidelberg (2004)

ALSOK (2009), `http://www.alsok.co.jp/` (retrieved 9, 2009)

Autodesk (2009), `http://usa.autodesk.com/` (retrieved 9, 2009)

Birk, A., Kenn, H.: RoboGuard, a teleoperated mobile security robot. Control Engineering Practice 10, 1259–1264 (2002)

Blender (2009), `http://www.blender.org/` (retrieved 9, 2009)

Browning, B., Tryzelaar, E.: ÜberSim: A Multi-Robot Simulator for Robot Soccer. In: Proceedings of the Second International Joint Conference on Autonomous Agents and Multiagent Systems (2003)

Calvo, R., Cordone, R.: A Heuristic Approach to the Overnight Security Service Problem. Computers and Operations Research (2003)

Carnegie, D.A., Prakash, A., Chitty, C., Guy, B.: A Human-Like Semi Autonomous Mobile Security Robot. In: 2nd International Conference on Autonomous Robots and Agents, Palmerston North (2004)

Chevaleyre, Y.: Theoretical Analysis of the Multi-agent Patrolling Problem. In: IEEE/WIC/ACM International Conference on Intelligent Agent Technology (2004)

Chien, T.L., Su, K.L., Guo, J.H.: The Multiple Interface Security Robot – WFSR-II. In: Proceedings of the 2005 IEEE International Workshop on Safety, Security and Rescue Robotics, Kobe (2005)

Dapper, F., Prestes, E., Idiart, M.A., Nedel, L.P.: Simulating Pedestrian Behavior with Potential Fields. Advances in Computer Graphics (2006)

Dr. Robot (2009), `http://www.drrobot.com/` (retrieved 9, 2009)

Everett, H.: A Computer Controlled Sentry Robot. Robotics Age (1982)

Everett, H., Gage, D.: A Third Generation Security Robot. In: SPIE Mobile Robot and Automated Vehicle Control Systems, Boston, pp. 20–21 (1996)

Faria, G., Prestes, E., Idiart, M.A., Romero, R.A.: Multi Robot System based on Boundary Value Problems. In: Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems (2006)

Gu, K.Y.: Motion Planning Method for Robotics Pavement Inspection. Master Thesis, National Taiwan University (2008)

Hazon, N., Kaminka, G.A.: Redundancy, Efficiency, and Robustness in Multi-robot Coverage. In: Proceeding of IEEE International Conference on Robotics and Automation, ICRA (2005)
Hung, W.H.: Patrol Path Planning Method for Security Robot. Master Thesis, National Taiwan University (2008)
Hung, W.H., Kang, S.C.: Physics-based Crane Model for The Simulation of Cooperative Erections. In: 9th International Conference on Construction Applications of Virtual Reality (2009)
Kimoto, K., Yuta, S.: Autonomous Mobile Robot Simulatorm - A Programming Tool for Sensor-Based Behavior. Autonomous Robots (1995)
Kuc, R., Siegel, M.: IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 766-778 (1987)
Kucuk, S., Bingul, Z.: An Off-Line Robot Simulation Toolbox. Computer Applications in Engineering Education (2009)
Laue, T., Spiess, K., Röfer, T.: SimRobot - A General Physical Robot Simulator and Its Application in RoboCup. In: Bredenfeld, A., Jacoff, A., Noda, I., Takahashi, Y. (eds.) RoboCup 2005. LNCS (LNAI), vol. 4020, pp. 173–183. Springer, Heidelberg (2006)
Lemoine, P., Pape, L.: Simulating Actions and Perception of Autonomous Mobile Robots in a Multi-agent Indoor Environment. Intelligent Robots and Systems, Intelligence for Mechanical Systems, Proceedings IROS 1991. IEEE/RSJ International Workshop (1991)
Luo, R.C., Chou, Y.T., Liao, C.T., Lai, C.C., Tsai, A.C.: NCCU Security Warrior: An Intelligent Security Robot System. In: The 33rd Annual Conference of the IEEE Industrial Electronics Society (IECON), Taipei (2007)
Luo, R.C., Su, K.L., Lin, K.C.: Overview of The Intelligent Security Robot. In: Chung Cheng, I. (ed.) IEEE Workshop on Advanced Robotics and its Social Impacts (2005)
Matsumoto, T., Miyazaki, T., Inaba, M., Inoue, H.: View Simulation System: A Mobile Robot Simulator using VR Technology. In: Proceedings of the 1999 IEE/RSJ International Conference on Intelligent Robots and Systems (1999)
Mckerrow, P.: Simulation of Sonar Echolocation. In: Proceedings First Workshop on Domestic Robots and Second Work Shop on Medical and Healthcare Robotics, IARP (1989)
Michel, O.: Webots: Professional Mobile Robot Simulation. Inernational Journal of Advanced Robotic Systems (2004)
Microsoft (2009), Microsoft Robotics Developer Center, `http://msdn.microsoft.com/en-us/robotics/default.aspx` (retrieved 9, 2009)
Microsoft (2009a), WindowsClient.NET, `http://windowsclient.net/` (retrieved 9, 2009)
Microsoft (2008), XNA. MSDN, `http://msdn.microsoft.com/en-us/xna/default.aspx` (retrieved 7-14, 2009)
MobileRobots (2009), `http://www.mobilerobots.com/` (retrieved 9, 2009)
Murphy, S., Wen, T.Y., Saridis, G.: Simulation of Cooperating Robot Manipulators on A Mobile Platform. IEEE Transactions Robotics and Automation (1991)
Neobotix (2009), `http://www.neobotix.de/en/` (retrieved 9, 2009)
Nvidia (2009), PhysX. Retrieved from Nvidia, `http://www.nvidia.com/object/nvidia_physx.html`

Oh, S.Y., Orin, D.: Dynamic computer simulation of multiple closed-chain robotic mechanisms. Robotics and Automation. In: Proceedings 1986 IEEE International Conference (1986)

Oliveira, H.d., Vasconcelos, G.: A Hybrid Search Method for The Vehicle Routing Problem. Annals of Operations Research (2008)

Pelechano, N., Allbeck, J., Badler, N.: Controlling Individual Agents in High-Density Crowd Simulation. Eurographics/ ACM SIGGRAPH Symposium on Computer Animation (2007)

ROBOWATCH (2009), http://www.robowatch.de (retrieved 9, 2009)

Martin, R.: Web Based Robot Simulation Using VRML. In: Proceedings of the 2000 Winter Simulation Conference (2000)

SANYO (2009), http://jp.sanyo.com/ (retrieved 9, 2009)

SECOM (2009), http://www.secom.co.jp/ (retrieved 9, 2009)

Shimosasa, Y., Kanemoto, J., Hakamada, K., Horii, H., Ariki, T., Suqawara, Y., et al.: Security service system using autonomous mobile robot. In: IEEE SMC 1999 Conference Proceedings of Systems, Man, and Cybernetics (1999)

SKS (2009), http://www.sks.com.tw/ (retrieved 9, 2009)

Steinfeld, A., Fong, T., Kaber, D., Lewis, M., Scholtz, J., Schultz, A., et al.: Common Metrics for Human-Robot Interaction. In: 2006 Human-Robot Interaction Conference (2006)

Su, K.L., Chien, T.L., Guo, J.H.: Design a Low Cost Security Robot Applying in Family. In: 2nd International Conference on Autonomous Robots and Agents, Palmerston North (2004)

Tarnoff, N., Jacoff, A., Lumia, R.: Graphical Simulation for Sensor Based Robot Programming. Journal of Intelligent and Robotic Systems (1992)

TMSUK (2009), http://www.tmsuk.co.jp/ (retrieved 9, 2009)

Turnell, D., Fatima, M., Turnell, Q.: SimBot - A Simulation Tool for Autonomous Robots. In: 2001 IEEE International Conference Systems, Man, and Cybernetics (2001)

Wang, J., Lewis, M., Gennari, J.: USAR: A Game-Based Simulation for Teleoperation. In: Proceedings of the 47th Annual Meeting of the Human Factors and Ergonomics Society (2003)

Evaluating the Usability of Virtual Environment by Employing Affective Measures

I. Rezazadeh[1,*], M. Firoozabadi[1], and X. Wang[2]

[1] Department of Biomedical Eng., Science and Research Branch, Islamic Azad University, Tehran, Iran
[2] Faculty of Built Environment- University of New South Wales – Sydney – Australia

Abstract. In this chapter a new approach based on exploring affective status and cues for evaluating the performance and designing quality of virtual environments have been proposed. Five individual experiments have been performed to analyse the effect of proposed affective computing approach in designing schemes. The results show that, by employing user's emotional states into the designing paradigms, the better results could be achieved.

Keywords: Affective Computing, Forehead Bioelectric signals, Brain waves, Virtual Reality, Affordance based designing.

1 Introduction

Today, Virtual Reality (VR) technology encompasses many fields from industrial to military, medicine, and entertainment applications. It should be noted that, when people are talking about Virtual Environment (VE) or VR, it could lead to misunderstanding with Computer Graphic (CG) simulations. CG is mostly concerned about the modeling, lighting and dynamics of environments and in general how to build a medium – so forth called Virtual World (Lee and Pakstas, 2002). But, designing a VE is something beyond designing a CG. When we become more curious about how people experience the simulated environment, behave within it and interact with its components, we are talking about VE or VR (Sherman and Craig, 2003).

VR should be displayed as much as realistic to convince its users of being immersed within it. Being immersed in the environment can be referred to mental or physical immersion and is a one of the key factor which determines the quality of VE, especially for training purposes. Mental immersion is the state of being deeply engaged (presence) within the environment and physical immersion is bodily entered into a medium. So, the concern could be how to design a VE which could bring the sense of immersion for its users to increase its performance (Sherman and Craig, 2003).

[*] Corresponding author.

X. Wang (Ed.): Mixed Reality and Human-Robot Interaction, ISCA 47, pp. 95–109.
springerlink.com

Let's have an overview on our activities of daily life (ADLs). We are surrounded by many devices and using them to accomplish our daily tasks. Mostly, we have experienced using a device which comes from an advanced technology, but there are some difficulties to cope with using it. After some trials, if we cannot handle the device, we reject to use it and we would like to use the simpler ones – maybe belongs to older days- .

The above fact is very important and determines the social acceptance of a technology or a device within the target community. This acceptance is critically important while the device is designed for training purposes. If the user could not keep up with using the device, then one can conclude that, the designer is not successful to achieve his goal. Now, one may ask *whether the device is really good?!* Here, "*good"* depicts two faces of a designing: *Affordances- Usefulness and Usability* (Pons et al, 2008 ; Zecca et al., 2002). Assuredly, most of the modern and advanced devices are designed to help people and facilitate their tasks, so they are useful. But most of them are designed based on engineer-centered approaches in which:

- The engineer (designer) thinks instead of the user
- The device is designed from its designer's point of view.
- The user MUST use the device if (s)he wants to get some benefits from it.
- Mostly, this type of designing burden physical fatigue and cognitive overload (or fatigue) to its user. Thus, the user wants to leave the using of the device

So, to cope with the above problems, the engineer can ask himself whether the device is usable and easy to be handled *by the user* or not. This way of thinking leads the *user-centered designing* approach in which:

- The main focus is on how to design a device which is more suitable and comfortable for the user.
- The user's feeling regarding the device is more important than the applied technology within the designed.

There are some ways for understanding the user's feeling and idea about the device during the usage period:

1. Using Questionnaire: One traditional way could be through using a questionnaire which some predefined questions are asked from the user. This way could reflect user's level of satisfaction regarding the device but has also some drawbacks like the reliability of the answers. For example, some users do not answer the questions properly because of some ethical or social issues.
2. Using Physical Performance Factors: The second way to understand the user's feeling is using some physical performance factors such as task completion time, achieve scores during the performance, and so on. These factors could mirror the level of expertise obtained during the performance.

However, the concern is whether this expertise factors could also reflect level of satisfaction during the performance. For example, it could be a case when the user burdens cognitive overload but still cope with experiment. Here the achieved scores might be high, but the cognitive overload could decrease the performance over long running period and increase the level of dissatisfaction. So, it is not a proper and efficient way to discover the user's feeling.

3. Using Affective Measures: The third way for exploring the user's satisfaction during the performance which is based on the user's affective status. In this approach the affective measures (emotional status) are extracted from the user during performance. Then, by comparing the affective measure to some standard and individual levels, the level of satisfaction could be discovered.

The following sections focus on affective measures extraction for monitoring the training performance in a virtual environment. Also, how to design a VE which could increase the sense of immersion and suits for the user without burdening extra cognitive load.

2 Affective Computing

One of the new approaches which has been gain many attentions in analyzing emotional feeling is affective computing. It means, computing that relates to, arises from, or deliberately influences emotions. This methodology could be employed for creating Human Computer Interfaces (HCIs) which have the ability to sense, recognize and understand human emotions, together with the skills. So, the final goal in affective computing is giving the ability to machines to respond intelligently to human emotion not creating emotions (Picard, 1997). Russell (1980) showed the human's emotional status could be indentified in the emotional space (Figure 1). Up to now, there are many studies in the recognition of emotional states. Recent studies show the benefits of different methods for extracting emotions using speech recognition facial expression, papillary dilation **and** also the physiological data (blood pressure, heart rate and temperature for example).

For example, Coronel et al. stated that facial muscle movements and facial muscle activities (fEMG) can be corresponding to certain facial expressions and are the most visual representation of a person's physical emotional states. Mahlke and Minge used emotional states which were extracted from fEMG to discriminate between usable and unusable computerized context by placing two pairs of electrodes on Zygomaticus major and Corrugators' supercili muscles to detect positive and negative emotional states, respectively. They concluded that the frowning activity is significantly higher in the unusable system condition than in the usable one. (Mohammad Rezazadeh, 2010).

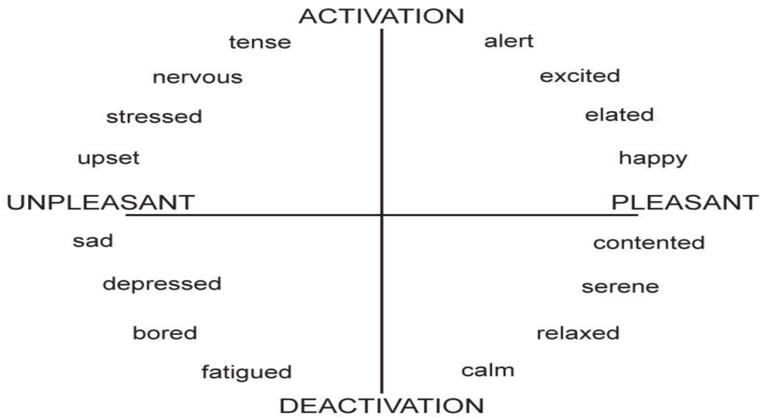


Fig. 1. Russell's emotional space shows different emotional status

3 The Affective Measures

As described in the section 2, there are many of methods for extracting affective measures from a human subject. Based on previous researches, one way is to employ the brain activity features (EEG) during the performance. It has been shown that, the characteristics of EEG signal and its complexity change due to different mental status. For example, the α sub-band activity (7 Hz- 12 Hz of EEG) is modulated by semantic memory processes and also related to attentional task demands and mental states monitoring (level of alertness, expectancy, mental relaxation and satisfaction for example) during the performance. Furthermore, it is believed that the α band represents oscillations of postsynaptic potentials in the neocortex, and is reduced in amplitude by moderate to difficult mental tasks. Several studies report a negative correlation between brain activity (mainly α band) under cognitive overload and performance indices. It is reported that, less complex EEG patterns have been observed in more intelligent individuals and in more creative ones. It was assumed that, the displayed reduction of the complexity of neural dynamics in high intelligent individuals is due to the inhibition of irrelevant and competitive activity. Alternatively, less intelligent individuals are characterized by more diffuse neural dynamics when performing the same task. One interpretation of such increases in amplitude in high intelligent individuals is a reduced neural network activity in regions not relevant for task performance – the neural efficiency hypothesis. In EEG signal, the measures of dimensional complexity reflect the complexity of neural generators or the relative number of concurrently oscillating neuronal assemblies and degrees of freedom in the competitive interaction between them (Doherty,.2002; Surdilovic and Zhang, 2006; Cyberlink, 2010).

4 Extracting Affective Measures from the Forehead

In this section, we employed multi-channel forehead bioelectric signal for collecting the bioelectric signals from the subject's forehead for extracting affective measure. As illustrated in Figure 2, three pairs of rounded pre-gelled Ag/AgCl electrodes were placed on the human subjects' facial muscles in a differentiation configuration to harness the highest amplitude signals (Mohammad Rezazadeh et al., 2009; Mohammad Rezazadeh et al., 2010a):

- One pair is placed on the subject's *Frontalis* muscle: above the eyebrows with 2cm inter-electrodes distance(Channel 2).
- Two pairs are placed on left and right *Temporalis* muscles (Channels 1 and 3)
- One ground electrode is placed on the bony part of the left wrist.

The Biopac system (MP100 model and ack100w software version) was used to acquire bioelectric-signals. It can accurately collect bioelectric-signals with the selected sampling frequency and store them in its own or PC’s memory. The sampling frequency and amplifier gain are selected at 1000 Hz and 5000, respectively. The low cut-off frequency of the filter is chosen to be 0.1 Hz to avoid motion artifact. In addition, a narrow band-stop filter (48Hz-52Hz) is used to eliminate line noise.

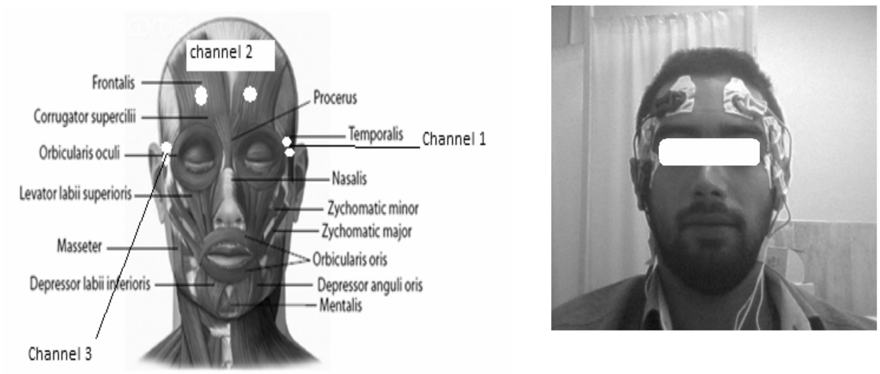


Fig. 2. Illustration of the electrodes configuration over Frontalis and Temporalis facial muscles (Mohammad Rezazadeh et al., 2009; Mohammad Rezazadeh et al., 2010a,b)

In our experiments the α sub-band of the forehead EEG (fEEG) was chosen because research has identified its relationship to different cognitive functions.

According to some studies, the *statistical entropy* could be employed as a measure of the system complexity and the degree of order/disorder of a signal. It could provide useful information about the underlying dynamical process associated with the signal. So, one could assume that, when the subject is satisfied with the environment while performing the requested task, the degree of disorder of the signals in α sub-band will be reduced []. So, the entropy measure could be considered as an affective measure (cue) which mirrors user's level of satisfaction during the task performance. The Log energy entropy measure (Formula 1) (*LogEn*, hereafter, entropy or statistical entropy) has been calculated for each given time slot using the same method as described in (Aydin, 2009).

??*H*??*LogEn*??= −??*i*=0??*N*−1??????(*log*??2????(*P*??*i*??(*x*)))??2???? (1)

5 Our Studies Based on the Forehead Affective Measures

Here, some of our previous studies regarding designing affective HMI are abstracted to show the potential capability of our proposed method for further applications.

5.1 Designing Adaptive Human Machine Interface Based on Affective Measures

The general block diagram of an affective controller is depicted in Figure 3. A human machine interface can be divided into physical layer, cognitive level, and its interface which controls the interaction process between human and machine. In the previous studies, the interface adapts itself with the physical performance indices eliciting from the physical level. But, the effect of the physical layer on the cognitive layer are mostly ignored or missed in the interface designing. The cognitive level could modify the physical level and the system's performance as well. So, the interface (control system) cannot cope with changes in the system, if

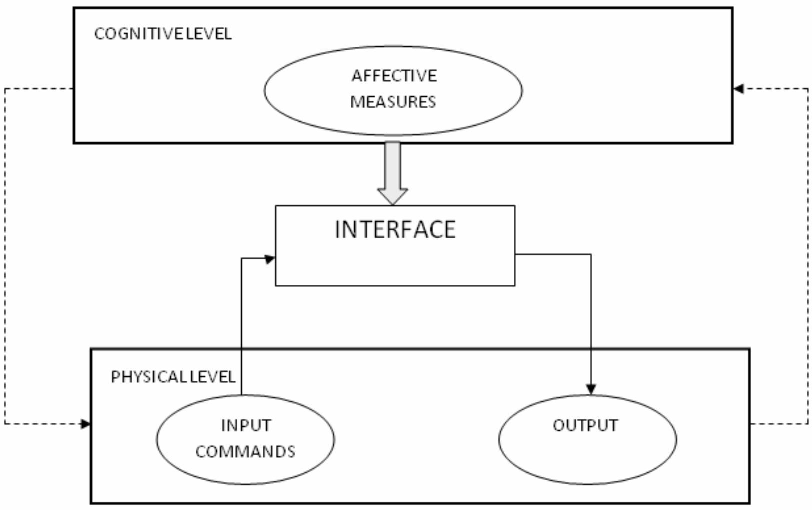


Fig. 3. The affective controller designing schematic. Dashed lines shows indirect influence of one part to another

it only be adapted to physical layer status, because the physical interactions within the interface have tight correlations with the affective status in the cognitive level.

Here, we propose that, the changes in cognitive level which are mirrored by affective measures could be considered as an important factor for modifying the interface. In the Sections 5.2 and 5.3 the described method has been employed as our approach to HMI designing for a virtual forearm prosthesis and crane, respectively (Mohammad Rezazadeh, 2010c).

5.2 A New Approach on Designing Affective Interface for Training Forearm Prosthesis Users in Virtual Environment

Fast and convenient adaptation of human assistive devices improves the quality of life particularly for amputees. This however, may not be achieved easily due to the variations of the users' physical and emotional status over time. In this study, for mitigating this problem, a collaborative and adaptable Human-Machine Interface (HMI) has been developed, which could adapt itself to the user's affective status and enhance the HMI's usability. The HMI was used to control a virtual forearm prosthesis in three different levels of difficulty for period of long run usage. The user's manipulation commands were recorded by placing two pairs of electrodes over Biceps and Triceps muscles. Physical performance measures for the requested tasks (expertise factor, EMG entropy and trajectory smoothness) were then calculated. The forehead bioelectric signals were recorded using one pair of electrodes placed on the subject's forehead for extracting the affective measure (the entropy of the alpha band of the forehead EEG) while performing the tasks. In this study the subject was asked to carry a virtual ball and put in the virtual basket using his muscles manipulation commands (Figure 4). By employing the described affective controller approach, the proposed HMI could adapt itself to the subject's affective status. The quantitative results of 7 subjects (including an

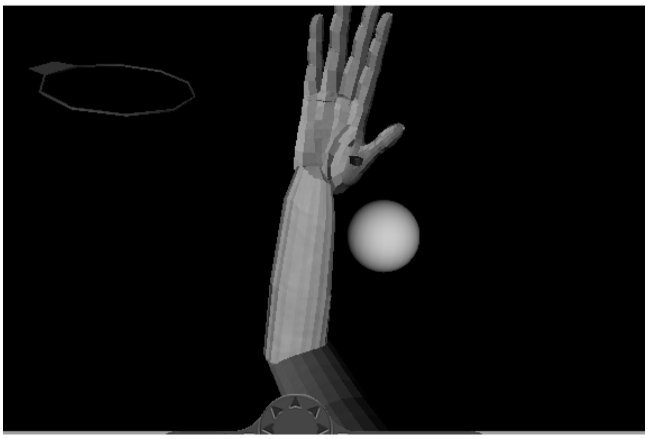


Fig. 4. The user is trying to get a virtual ball and then put it into the basket using his Biceps and Triceps muscles.

amputee) show that, the test group who used the proposed affective HMI achieved better physical performance measures in comparison with those in the control group who have not used the affective method. Furthermore, according to the results of questionnaires, the level of satisfaction within the test group was greater (Mohammad Rezazadeh, 2010c).

5.3 Using Affective Human-Machine Interface to Increase the Operation Performance in Virtual Construction Crane Training System

In the construction industry, some progresses have been achieved by researchers to design and implement environments for task training using VR technology and its derivatives such as Augmented and Mixed Reality. Although, these developments have been well recognized at the application level, however crucial to the virtual training system is the effective and reliable measurement of training performance of the particular skill and handling the experiment for long-run. It is known that motor skills cannot be measured directly, but only inferred by observing behaviour or performance measures. The typical way of measuring performance is through measuring performance time and accuracy (physical performance indices), but can be supported by indirect measurement of some other factors. In this study, a virtual crane training system has been developed which can be controlled using control commands (Table 1) extracted from facial gestures via channels 1,2 and 3 (Figure 2) and is capable of lifting up load/materials in the virtual construction sites (Figure 5).

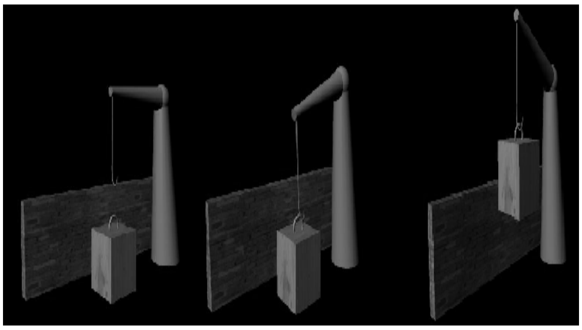


Fig. 5. Three stages of manipulating the virtual crane (Lifting, handling and releasing the load)

Then, we integrate affective controller concept into the conventional VR training platform for measuring the cognitive pressure and level of satisfaction during performance using human's forehead bioelectric signals (Channel 2). By employing our novel control scheme, the designed interface could be adapted to user's affective status during the performance in real-time. This adaptable user interface approach helps the trainee to cope with the training for long-run performance, leads to 20 % increasing performance factor and more effective transfer of learning to other environments (Mohammad Rezazadeh et. al, 2010d).

Table 1. Facial manipulating commands and their corresponding channels

Gesture No.	Gesture Name	Related Command	Most Relative Physical Data Channel
1	Smiling	Move Forward	Channel 1 & 3
2	Pulling up right lip corner	Move Right	Channel 3
3	Pulling up left lip corner	Move Left	Channel 1
4	Opening mouth (like to say 'a' in 'apple')	Move Backward	Channel 1 & 3
5	Clenching Molar teeth	Lift/Release the load	Channel 1 & 3

5.4 Relation between Cognitive Load and Affective Measures

To study the effect of cognitive load over the subject's affective status the following study has been performed.

Twelve healthy volunteers aged 20 ±2 participated in this experiment and divided equally into the test and the control groups. For the test group each subject wear HMD from *iglass*™ during the experiment and for the control group the game scene was shown through a laptop monitor. Then each subject was asked to play **Svetlograd**™ **1.0** – Fresh color-matching shooter- (Figure 6.1) using computer's mouse in two different levels of difficulty- Easy and Hard – each for 5 minutes (Figure 6.2).

The obtained scores during the performance are considered as physical performance and shown in the Table 2. It should be noted, because the subjects gained more expertise as time passed, the average obtained scores increased over

Fig. 6.1. A snap-shot for **Svetlograd 1.0**

Fig. 6.2. A subject while playing **Svetlograd 1.0**

Table 2. The physical performance measure of the test and control group (P<0.01)

Difficulty level ***The average Obtained Score***	***Easy***	***Hard***
Control group	*21170*	***54320***
Test group	*25308*	***117250***

the time. But despite the effect of learning transfer, it is clear that the obtained results in the test group are much greater respect to the control group, especially in the more difficult level of the game, because the HMD provided the subjects more concentration.

Now, to find out whether there is any relationship between cognitive load and affective measures, the alpha sub-bands of EEG in the control and test groups signals were captured and the *LogEn* features were extracted. According to Table 3, the results showed that for the test group the entropy of EEG is lower respect to control group in both easy and difficult levels. It is clear that the entropy level increased when the game moved from easy to hard level. The questionnaires showed that, after time passed in the hard level, playing the game was being more difficult for the subject, but he still want to keep up with experiment. The achieved affective measures could show these changes in emotional states. Also, as shown in Table 3, the increasing rate of entropy from the easy to the hard level in the test group is 11% lower respect to the control group ($p<0.01$). The reason

Table 3. The affective measure of the test and control group (P<0.01)

Difficulty level	***Easy***	***Hard***
The average affective measure		
Control group	*0.052*	***0.627***
Test group	*0.031*	***0.0348***

could be using HMD which brings more comfort for the user during the performance and in one word provides more immersion (Khanjan nejad and Forghani, 2010) .

5.5 How Does VRE Designing Effect on User's Performance?

As mentioned above, one of the main concerns on designing a VRE is to know whether the designed environment is good enough for the training purposes. Many studies show that the training environment should have some correlation with the requested task from the trainee to perform. But up to now, there are few studies which compare the effect of relevant and irrelevant training environment on the total performance of the system.

Here, for evaluating the effect of VRE's on the user's performance measures, we have designed a virtual face (Figure 7) which responses intelligently to the user's emotional status eliciting from the user's brain activities. If the subject can

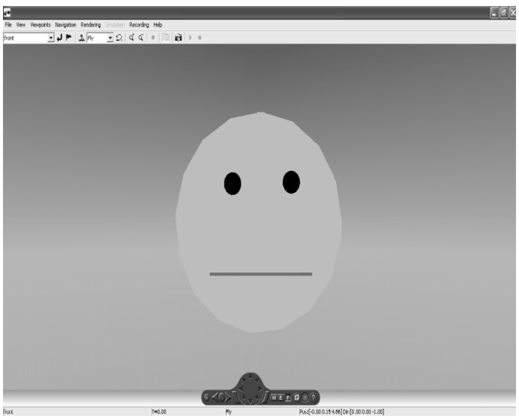


Fig. 7. The virtual face designed to evaluate the different designing approaches

retain or increase his/her level of satisfaction (increase the amplitude of α wave from Channel 2) then the user gain positives score and vice versa.

Here, there are two different scenarios which the face response to achieved scores. In the first scenario, the virtual face smiles and being happier when the users gains continuous positive scores. But, in the second scenario, the positive scores lead to a sad virtual face (Figure 8).

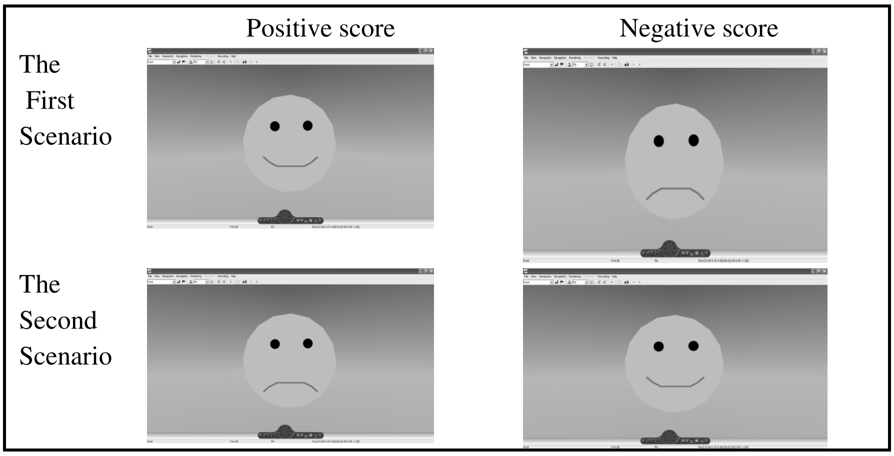


Fig. 8. Two different scenarios for evaluating the effect of VRE

Six subjects participated in the above experiment. They were asked to gain more scores by looking at the virtual face. For the group who employed the smiling virtual face, they achieved 37% more scores in comparison with the user of sad virtual face within 10 minute experimental period. Also, the questionnaires showed that happy virtual face users were relaxed with the experiment.

The α brain wave amplitude increases when the subject is relax and satisfied about the context. So, the first scenario (happy virtual face) has more correlation with the requested task and that is why the subjects obtained better results (Tajziyechi, 2010).

5.6 Identifying the Effect of Music Playing on the Affective Status

It is a very old belief that music is a link between cognition and emotion and that music can influence autonomous neural system reactions both in an arousing and a calming fashion.

One of the approaches in assessing human emotions is represented during music performance. In this study two types of music which was selected based on Russell’s emotion space (Figure 1) - negatively excited and calm/pleasant states- were discriminated using the affective measures described above eliciting from Channel 2.

Ten healthy female volunteers aged between 19 to 24 (mean 21.6) with approximately same level of IQ participated in experiment. None of the participant did play any musical instruments before and also they did not have any hearing problems. Each subject was asked to close her eyes and be relaxed. Then her relax signal was recorder simultaneously. Two types of music were played for her with 5 minutes gap time.

Three states of relaxed, calm-pleasant and negatively-excited signals were compared with each other using the Wilcoxon statistical test. This test was applied over affective measures from α and β brain waves, and facial EMG to see if whether there was any significant statistical different among them.

Table 4 shows that the above states could be separated statistically. We conclude that when designing a virtual environment for training or entertainment purposes the played music during the performance is a key factor which can affect on the overall user's emotional status. By referring to Figure 3, more attention should be taken to the type of music during the performance.

Table 4. The statistical analysis of three states: relaxed, calm-pleasant and negatively-excited

		P_{value} Rest with Calm-Pleasant	P_{value} Rest with Negatively-Excited	P_{value} Calm-Pleasant with Negatively-Excited
Alpha-band	*Channel 1*	0.0280	0.0130	0.2410
	Channel 2	0.0590(boarder)	0.0090	0.0370
	Channel 3	0.0220	0.0280	0.0030
EMG-band	*Channel 1*	0.0070	0.0130	0.0930
	Channel 2	0.0590(Boarder)	0.0090	0.0740
	Channel 3	0.0130	0.0050	0.114
Beta-band	*Channel 1*	0.2400	0.0160	0.0920
	Channel 2	0.6460	0.0120	0.0210
	Channel 3	0.0640	0.0070	0.0610

6 Conclusion

Our proposed methodology for designing a VE designing scheme is depicted in Figure 9. It consists of some building blocks where the user is located at the center. A good designing should comply with the user's needs and bring satisfaction for the user. So, studying these factors is important for the system's designer. Also, when monitoring the system performance, the physical performance indices cannot simply illustrate the user's comfort with the systems. The self-report data and affective measures should be also collected for achieving better understanding regarding the system's status.

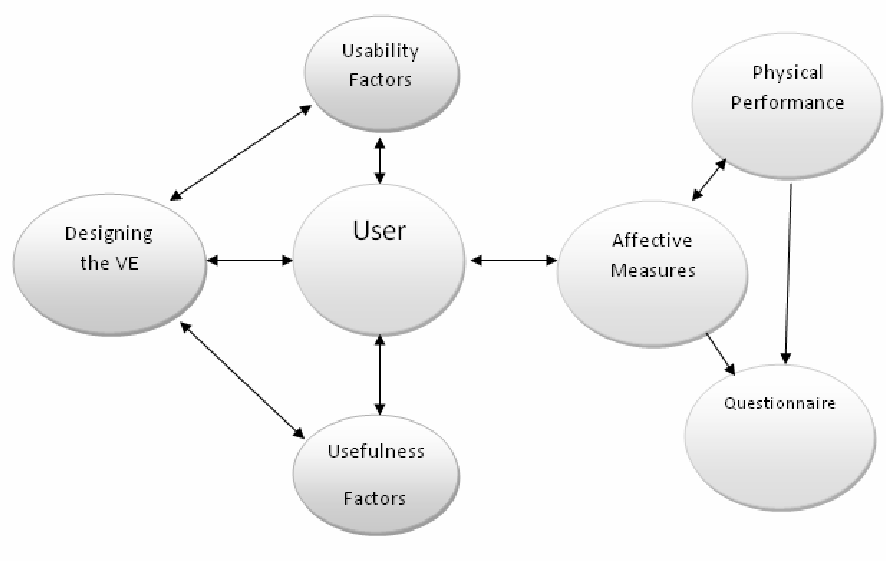


Fig. 9. The proposed VE designing scheme

References

Aydin, S., Saraoglu, H.M., Kara, S.: Log energy entropy-based EEG classification with multilayer neural networks in seizure. Annals of Biomedical Engineering 37(12), 2626–2630 (2009)

Cyberlink (2010), `http://www.brainfinger.com` (last accessed on May 19, 2010)

Doherty, E.P., Cockton, G., Rizzo, J., Blondina, B., Davis, B.: Yes/No or Maybe—further evaluation of an interface for brain-injured individuals. Interacting with Computers 14(4), 341–358 (2002)

Hashemi Rad, R.: Identifying the effect of music playing on the affective status. BSc thesis, Department of Biomedical Eng., Science and Research branch of Islamic Azad University (2010)

Khanjannejad, C., Forghani, H.: Investigating relationship between cognitive load and affective measures. BSc thesis, Department of Biomedical Eng., Science and Research branch of Islamic Azad University (2010)

Lee, E., Pakstas, A.: Over view of computer-based virtual reality. In: Pakstas, A., Komiya, R. (eds.) Virtual reality technologies for future telecommunication systems, pp. 3–7. Wiley, UK (2002)

Mohammad Rezazadeh, I., Wang, X., Wang, R., Firoozabadi, S.M.P.: Toward affective handsfree human-machine interface approach in virtual environment-based equipment operation training. In: The 9th Int. Conference on Construction Applications of Virtual Reality, Sydney, Australia (2009)
Mohammad Rezazadeh, I., Firoozabadi, S.M.P., Hu, H., Hashemi Golpayegani, S.M.P.: Determining the surface electrodes locations to capture facial bioelectric signals. Iranian J. Med. Physics (in press, 2010a)
Mohammad Rezazadeh, I., Firoozabadi, S.M.P., Hu, H.: A novel human-machineiInterface based on recognition of multi-channel facial bioelectric signals. International Journal of Human Computer Studies (2010b) (submitted)
Mohammad Rezazadeh, I., Firoozabadi, S.M.P., Hu, H., Hashemi Golpayegani, S.M.P.: Affective Human-Machine Interface for Improving Training Performances of Virtual Myoelectric Forearm Prosthesis. IEEE Trans. on Affective Computing (2010c) (submitted)
Mohammad Rezazadeh, I., Wang, X., Firoozabadi, S.M.P.: Using affective human-machine interface to increase the operation performance in virtual cinstruction crane training system: A novel approach. Elsevier J. Automation in Construction (2010d) (submitted)
Picard, R.: Affective computing. MIT Press, Cambridge (1997)
Pons, J.L., Ceres, R., Calderon, L.: Introduction to wearable robotics. In: Pons, J.L. (ed.) Wearable robots: Biomechatronic Exoskeleton, pp. 1–16. Wiley & Sons, UK (2008)
Russell, J.A.: A circumplex model of affect. Journal of Personality and Social Psychology (1980)
Sherman, W.R., Craig, A.B.: Understanding virtual reality: interface, application, and design, pp. 2–38. Elsevier, USA (2003)
Surdilovic, T., Zhang, J.Q.: Convenient intelligent cursor control web systems for internet users with severe motor-impairments. International Journal of Medical Informatics 75(1), 86–100 (2006)
Tajziyechi, M.: How does virtual reality environment effect on user's performance?, BSc thesis, Department of Biomedical Eng., Science and Research branch of Islamic Azad University (2010)
Zecca, M., Micera, S., Carrozza, M.C., Dario, P.: On the control of multifunctional prosthetic hands by processing the electromyographic signal. Critical Reviews in Biomedical Engineering 30(04-06), 416–419 (2002)

A Mixed Reality Based Teleoperation Interface for Mobile Robot

X. Wang and J. Zhu

The Faculty of Built Environment, The University of New South Wales, Australia

Abstract. The human-robot interface system is the key to extending the application field for next generation robot systems. Conventional interface for robots have been widely used to perform tasks such as reconnaissance, surveillance and target acquisition. However, it is too complex and difficult to use for people who do not have sufficient situation awareness. Additionally, constructing mental models of remote environments is known to be difficult for human operators. This paper proposes a mixed reality interface for remote robot using both real and virtual data acquired by a mobile equipped with an omnidirectional camera and a laser scanner. The MR interface can enhance the current remote robot teleoperation visual interface by combining real environment and virtual information together on a single display to efficiently improve situation awareness, to facilitate the understanding of surrounding environment, and to predict the future status. The computational model that describes the triangle relationship among the mobile robot, the operation and intelligent environment also be discussed in this paper.

Keywords: Mixed Reality, Situation awareness, Teleoperation, Remote robot.

1 Introduction

Robots are typically used to perform tasks such as reconnaissance, surveillance and target acquisition, which traditionally human access is impractical. There is real case of robots being used at disaster searching. The incident of World Trade Center Towers require the robot equipment to enter miniature voids and areas where is too dangerous for human. (see Fig. 1).

CRASA teams (Fig .2) deployed eight times to the rubble pile during the rescue task. (Casper, 2003). The focus of the task was the use of robots to search for victims and exams voids that could not be reached by human.

Casper and Murphy (2003) evaluate collected data and results from the task of the robot and they argue that robot play critical role in the rescue task. The reason is multifold. Firstly, because of the limitation of the size, extreme heat, or toxicity of the environment, robot can access places but human can not. Fig.3 shows a good example. The white square marks the entrance to the void that could not be

X. Wang (Ed.): Mixed Reality and Human-Robot Interaction, ISCA 47, pp. 77–93.
springerlink.com

Fig. 1. Robot used in damaged buildings surrounding the main rubble pile at Ground Zero (Casper, 2003)

examined by human (Casper, 2003). Secondly, rescuing safety and effectiveness is serious issue and that robots are expendable. Finally, robot can significantly reduce the secure time consuming and aid in medical support (Drury, 2001).

However, current mobile robot technology is not well developed for rescue robots and regardless, such robots are fully autonomous or telerobotic. Therefore human–robot interaction is a key component of a successful rescue system. Casper and Murphy's (2003) analysis of video data collected during the World Trade Center disaster response found that a variety of human–robot interaction issues impacted performance of human–robot teams on the pile. Operator's lack of awareness regarding the state of the robot and regarding situatedness of the robot in the rubble is the most relevant factor to this study (Burke et al, 2004). Operators also had difficulty in linking current information obtained from the robot to existing knowledge or experience (Casper, 2003). The Florida task force and World Trade Center human–robot interaction studies reveal difficulties in operator teleproprioception and telekinesthesis, consistent with the problems described in Sheridan (1992).

Fig. 2. (Top) the Robot and (bottom) operator control unit (Casper, 2003)

Fig. 3. The white square marks the entrance to a void searched, approximately 0.3 _ 0.5 m cross section (Casper, 2003)

Basically, these problems occur with the situation that the robot operator is distant from the actual robot based on such settings limitations. In order to operate a robot efficiently at remote spaces, it is important for the operator to be aware of the environment around the robot so that the operator can give informed, accurate instructions to the robot. This awareness of the environment is often referred to as situation awareness.

2 Conventional Ways Overview

It had previously noted by Drury (2001) that most problems encountered when navigating robots have resulted from the humans' lack of awareness of the robot's location, surroundings or status. Situational awareness is defined by Endsley (1988) as "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future." Yanco et al (2005) modified the definition for HRI, giving a definition of situation awareness as the perception of the robot's location, surroundings, and status; the comprehension of their meaning; and the projection of how the robot will be have in the near future.

In order for a human operator to control a robot effectively, it requires him or her understanding of the remote environment and situation around the robot. It means he or she need to keep the high level of the situational awareness. Since the robot is in a remote distant from the human operator and cannot be directly observed, the necessary information for the human operator to develop an understanding or awareness of the robot's situation comes from the user interface. The usefulness of the human robot interface depends on the manner in which the information from the remote environment is presented.

Despite the importance of situation awareness in remote-robot operations, experience has shown that typically interfaces between human and robots do not sufficiently support the operator's awareness of the robot's location and surroundings. The case of World Trade Center is a good example. According the research by Casper and Murphy's (2003), the robots were useful because they were able to get into small, dangerous areas that were inaccessible for rescuing workers; however, it was quite difficult for the operator to navigate the robot while searching the environment because the robots only provided video information to the operator. Woods (2004) believed that there is a limitation of the robot which comes from the limitation of views of most cameras creates a sense of trying to understand the environment through a 'soda straw' or a 'keyhole'. It makes difficult for an operator to be aware of the distance between robot and obstacles.

There are different methods that could improve the situation awareness in telerobotics. However, experience has shown that operators typically do not demonstrate sufficient awareness of the robot's location and surroundings. (Nielsen et al, 2007). According to the research by Woods (2004), if the operator was provided video information only, it would create a sense of trying to understand the environment through a "soda straw" or a "keyhole". Alfano and Michel (1990) believe that the limited view of the robot's environment makes it difficult for an operator to be aware of the robot's proximity to obstacles. Experiments by Yanco et al (2005) revealed that even more sensors were used and operators were more familiar with the robots, the result was not as good as expected.

Conventional interfaces are one possible reason that operators demonstrated poor situation awareness in the previous studies. For instance, conventional 2D interfaces present related pieces of information in separate parts of the display. According to the research by Endsley (1988), Lee (2001) and Scholtz (2002), such 2D interfaces require the operator to mentally correlate the sets of information, which can result in increased workload, decreased situation awareness and performance. Ricks et al. (2004) argued that these negative consequences arise from a cognitive perspective because the operator has to perform mental rotations between different frames of reference frequently (e.g., side views, map views, perspective views) and fuse information. To improve situation awareness in human–robot systems, Yanco et al. (2005) recommended: (1) using a map; (2) fusing sensor information; (3) minimizing the use of multiple windows; and (4) providing more spatial information to the operator. These recommendations are consistent with observations and recommendations from other researchers that involve human–robot interactions (Casper, 2003. Burke, 2004. Murphy and Rogers, 2003).

Many robots used for studying human-robot interactions (HRI) have multiple sensors and a map-building algorithm in addition to the camera in comparison to the robots used at the World Trade Center. However, despite better equipment, recent experiments suggest that operators are still experiencing inadequate levels of situation awareness. The experiment by Yanco et al (2005) had searched a mock environment looking for victims using a robot that had sonar, laser, camera, and map-building capabilities. They found that even operators spend almost 30% of their time to acquire situation awareness; the operators often expressed and demonstrated confusion concerning the robot's location relative to obstacles, landmarks, and previous locations. A major complaint from the operators was that map built by the robot was totally useless because it did not help then to understand the robot's location.

There is another field study of rescue robots which was involved in the Urban Search and Rescue (USAR) training exercise. Burke et al. (2004) found that operators spent almost 54% of their time acquiring situation awareness as opposed to navigating the robot. Similarly, despite spending most of their time acquiring information about the robot and the environment, the participants still had difficulty using the robot's information to improve their own understanding of the search and rescue site.

The lack of situation awareness observed in the previous examples is not only a problem in rescuing personnel and others who may be unfamiliar with the robot and the interface. Similar results of poor situation awareness were found in a 2001 AAAI1 USAR competition(Drury, 2003). In this competition, it was the engineers that developed the robots and the interface, who competed using their own equipment. Despite the operator's familiarity with the equipment and the abundance of information (laser, sonar, map, and camera), the operators still demonstrated a lack of awareness of the robot's location and surroundings by bumping into obstacles and even leaving the experiment arena (Drury. 2003).

Operators usually experience low level of awareness in the previous studies because 2D (conventional) interfaces were used to display information from the robot to the operator (Nielsen, 2004). Conventional interfaces could make it not that easy for an operator to maintain the awareness situation of the robot's because some of related information are presented in different parts of the display. There is one essential problem that the definitions for telepresence are focusing on the accuracy where an environment is presented instead of focusing on communicating effective environmental cues. It has result in the use of displays such as those shown in Fig. 4, which illustrate accurate information from the environment (Ricks, 2007). However the information is presented in a diffused way rather than integrated manner. The operators have to mentally combine the data into a holistic representation of the environment because of those disparate information.

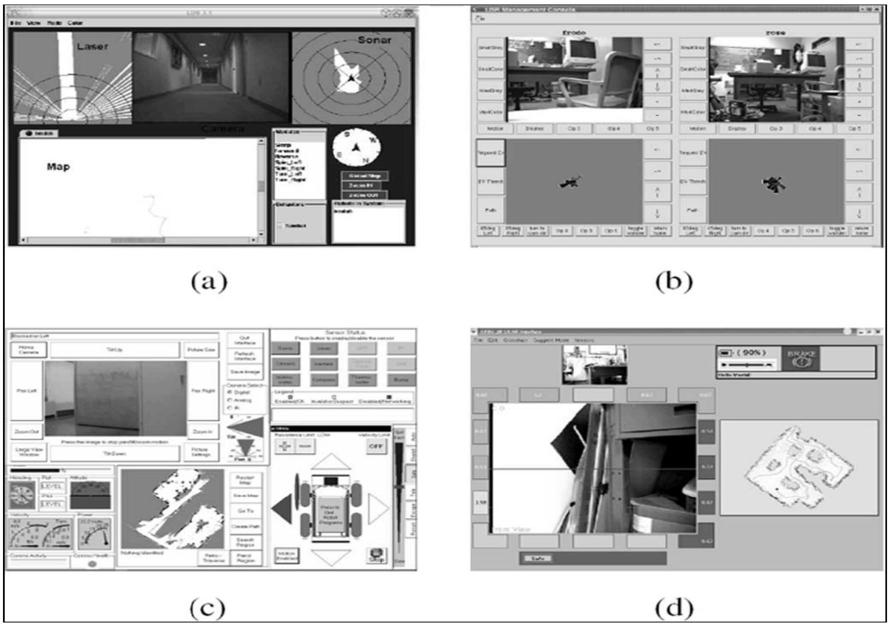


Fig. 4. Interfaces in the standard paradigm present information in separate windows within the display. (Ricks, 2007)

When related information is presented in different places, an operator must mentally correlate the sets of information, which can result in decreased situation awareness and decreased performance. Therefore, to improve situation awareness and performance, the ideal interface should correlate and present related information in a single part of the interface, thereby reducing the operator's cognitive workload required interpreting the information.

Accuracy is critical factor when most conventional interfaces used to interact with remote robots. However, communicating effective environment cues do not mentioned when information is presented to the operator (Nielsen, 2006).This factor has result in the use of displays such as those shown in Figure 4, which show information from the environment, but present it in multiple windows on the display. Nielsen (2006) believes the operator has to mentally combine the data into a cognitive map of the environment because of the disparate information. This approach of interface design follows the constructivist theory of perception, which claims that smaller, individual elements are combined to give perceptions (Helmholtz, 1962).

Bruemmer et al. (2002) developed a robot system for remote operations that uses behavior-based algorithms to create a mixed-initiative human-robot team. Information is displayed to the operator via a typical 2D display as shown in Figure 4. Baker et al. (2007) have simplified the interface designed by Bruemmer et al. (2002) in an effort to improve the human-robot interactions as shown in Figure 2.1(d). Both interfaces present information using a conventional robot-centric approach with separate windows for different information sources. Typically these interfaces are correspondingly useful for robot system development and testing as they provide enough information for the engineer to diagnose any problems that exist, but they may not adequately support an operator's situation awareness in many interesting remote environments.

One of the disadvantages when navigating a robot with a conventional interface is that typical cameras have a correspondingly narrow FOV. For example, a human's lateral field of view is normally 210 degrees; in contrast, the camera on robot usually has a field of view of 37 degrees only. The field of view that the operator has in an environment is very important to navigation. A poor condition of field of view has been attributed to negatively affect locomotion, spatial awareness, and perceptions of self-location. Further, Woods described using video to navigate a robot as attempting to drive while looking through a 'soda straw'. Operator typically does not have a good sense of what is to the 'sides' or 'shoulders' of the robot it the one the main challenges of the teleoperating and obstacles that need the most attention are typically outside of the field of view of the robot (Woods, 2004).

One method for overcoming a narrow field of view is to use multiple cameras. For example, Hughes et al. (2003) used two cameras and proved that it improved an operator's ability to perform a search task. Nielsen (2004) believe there is another method for improving field of view is to use a panospheric camera, which gives a view of the entire region around the robot. Those approaches may help operators better understand what is all around the robot, however it require fast communications to send large or multiple image with minimal delay and it also clearly increase the cognitive work burden of the human operators. Therefore the new system restricts attention to robots with a single camera.

Another method for acquiring situational awareness would be using of the map generated by the robot system and displayed on the system's interface. However, the subjects noted that the map was not that useful as expected. Drury (2005) believes that this problem occurred due to the fact that the depiction of the robot on the map was a small dot, while walls were also marked using pixels. If the robot was moving, people could see the dot representing the robot move around the screen. However, since the map is presented diagonally to the right, below the video screen, the operator could not watch the map updating while concentrating on the video screen to drive the robot.

3 Mixed Reality Interface

There is another method to improve robot teleoperation: using virtual environments to create a virtual scene that represents the real environment. According to the research of Wang and Dunston (2006), one of the major issues involved is the situation awareness is that the human operator has to keep during teleoperation task. The operator can be assisted by Mixed Reality (MR) to maintain a high situation awareness. Several prototype-phase research projects have been implemented to apply Augmented Reality (a sub-mode of Mixed Reality) technique into various teleoperation systems (Gu et al. 2002; Lawson et al. 2002; Fuchs et al. 2002). All their work focused on applying Augmented Reality in teleoperation and ignored other sub-modes of Mixed Reality and Augmented Virtuality. The proposed main goal of this article is to apply Mixed Reality technique in setting up a natural interface where surrounding scene visualization and structure are more available as a spatial reference aid.

Wang and Kwork (2007) have proposed an Augmented Reality-based autonomous robotic inspection system for interior built environment by combining the strength of by mobile robots and Augmented Reality. A mobile robot is equipped with a camera where the captured images are used to build an Augmented Reality system. Operators may teleoperate the robot and are provided with enriched information from AR for efficient inspections.

In this research project, a Mixed Reality-based human-robot visual interface is conceptualize, designed, developed and evaluated as an approach to improve operators' awareness of a remote mobile robot based on the above recommendations. The MR interface is based on Gibson's (1979) affordances theory, which claimed information that acts appropriately is inherent in the environment. Applying this theory to remote robots means that an operator's decisions should be made based on the operator's perception of the robot's affordances in the remote environment.

The notion of effective information presentation and ability to act on information is also addressed by Endsley's (1988) definition of situation awareness, as well as Zahorik and Jenison's (1998) definition of telepresence. The MR interface uses multiple sets of information sources from the remote robot to re-construct a 3D virtual environment (counterpart of the real remote environment) that is augmented with real video information and virtual video information. This MR representation of the remote environment combines real video, map, and virtual robot avatar into one single integrated view that is presented to the operator. The 3D MR interface is used to support the understanding of the relationships between the different sets of information. This approach presents the environment's navigational affordances to the operator and visualizes how they are related to the robot's current position and orientation.

MR merges the real and virtual objects and the development methodology is based on this concept. Figure 6 (Wang, 2006) shows three scene worlds: inner virtual world, real world, and outer virtual world. The inner virtual world defines the parameters of intrinsic virtual camera, as well as other spatial parameters and the virtual entities in AR portion are rendered with those parameters. Real world refers to the real-time 3D video, and the virtual entities in the intrinsic virtual world are registered in that video. The real cameras mounted on the remote machine determine the perspective of real world video. In the AR portion, virtual entities and the video that is realized by equating the parameters are seamlessly aligned by these two cameras. The outer virtual world is rendered based on sampled data under the outer virtual camera. In the AV portion analogously, when the parameter of outer virtual camera is set as the same to the parameter of the real video, then the content in the video could be seamlessly integrated with the virtual surrounding. It means perspective changes triggered by manipulating the real camera could induce the corresponding perspective changes in the outer virtual environment. Setting the parameters of these three cameras could create a seamlessly mixed scene under such circumstance, the geometry and position of all the entities (either real or virtual) based on the same global reference system. The poses of the extrinsic and intrinsic virtual cameras are then synchronized to the pose of the real camera.

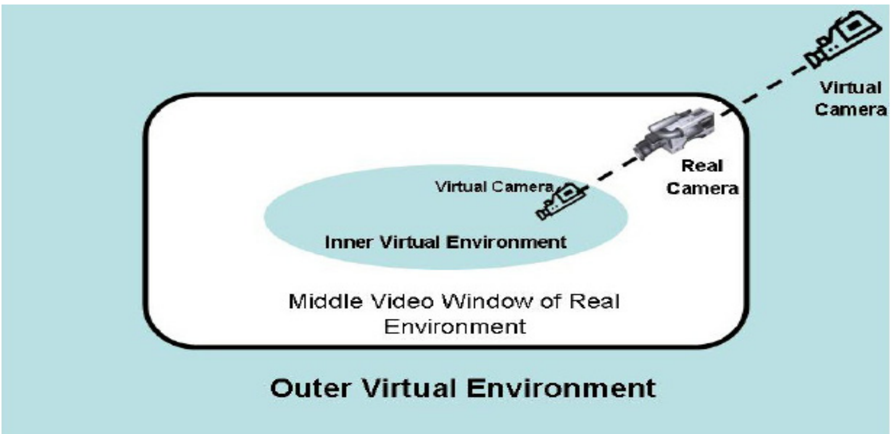


Fig. 5. Definition of mixed scene (Wang and Dunston, 2006)

When camera navigates around the environment, it acquires images of the world and the controlling process that is capable of extracting video stream from the camera signal. It is acquired that the camera parameters and video stream should be automatically and continuously updated in a real time manner. Once the virtual camera' parameters are equated to that of real camera, it will then project the relevant virtual objects geometries onto extrinsic and intrinsic virtual cameras' image planes that respectively match the real camera in the inner and outer virtual world. With the help of this 3D to 2D projection, it is possible to extract the relevant portion of the real camera image could as a video stream and then used in the virtual world. Because the camera moves around in the physical environment, the virtual world could come to resemble more closely the modeled real surroundings. The extent of seamlessness of border between the video and the outer virtual environment could be decided by the following factors: calibration, accuracy of modeling of real surroundings, lighting and the angle which the camera makes with the objects surface.

Milgram and Drascic (1994) defined Mixed Reality displays as “particular subset of Virtual Reality (VR) related technologies that involve the merging of real and virtual worlds somewhere along the ‘virtuality continuum’ ”. On one side of the virtuality continuum are real environments, on the opposite side are virtual environments. Mixed reality, sometimes referred to as augmented reality, is the domain between the two extremes. Milgram and Drascic (1994) point out that most of the work in mixed reality has been done using head-mounted displays (HMDs) that could either provide video feedback of the real world or allow the user to obtain some direct visibility of the real world.

Milgram and Kishino (1994) present a taxonomy that addresses the real and virtual aspects of mixed reality environments. The dimensions of their taxonomy include: extent of world knowledge, reproduction fidelity, and extent of presence metaphor. This taxonomy fits well with the needs of this research because it could separate the notion of presence into “image quality” and “immersion”, in contrast to other taxonomies whose primary goal is creating an immersed presence. This separated notion of presence provides a useful category for monitor-based mixed-reality displays, which are the tools of this research project.

Drascic and Milgram (1993) have discovered that the use of a stereoscopic display improves the user’s interaction because it presents depth information to the user directly. In contrast, monoscopic video images require the operator to interpret shadows and reflections to infer spatial relations. Milgram and Drascic (1994) discussed the use of augmented reality as a means to overlay a stereoscopic display with virtual information to facilitate connections between a human and a robotic arm. Their approach is based on the ARGOS (Augmented Reality through Graphic Overlays on Stereovideo) system and combines elements such as a virtual pointer, a virtual tape measure, and virtual landmarks to help the user control the robotic arm. By gathering stereoscopic information from the remote environment, the user is able to view a virtual 3D scene of the environment. Then, the user could work within the augmented reality environment to determine exactly how they want to manipulate the robotic arm. Once the commands are decided, the user sends them to the remote robot and the commands are actuated. This approach is similar to the interactions with the VEVI control system.

Meier et al. (1999) explored the possibility of using sensor fusion to make the operator be more aware of the environment around a mobile robot. Their display is typical of sensor fusion approaches for mobile robotics in which they overlay real video information with depth and other virtual information .By this particular approach the video is from a stereoscopic camera and it is combined with sonar range information to create a colored depth map. Moreover, the image displays a projected grid, which is overlaid on the ground and obstructed by aboveground obstacles. The grid cells are of similar sizes of the robot to support the operator’s comprehension of distances. The problem with most of these sensor fusion based displays is that even though the video is augmented with virtual information, the field of view of the environment is still limited by the field of view of the camera.

There is another example of a mixed-reality display. Johnson et al. created an “EgoSphere” (a term first proposed by Albus) to enhance their robot interface (2002). The EgoSphere consists of a 3D sphere around the robot on which interesting observations are portrayed. However, they did not find the EgoSphere to be particularly useful with a mobile robot. It is argued that an EgoSphere is probably more appropriate for an augmented reality display where the operator is wearing a head-mounted display.

Suomela et al. (2003) have developed a fully adjustable three-dimensional map that supports traditional two-dimensional map views and a full range of perspective views for a head-mounted display. They found that a single perspective view is useful sometimes, but different participants preferred different perspectives. Further, they identified situations where a “north-up” map is better than a “north-forward” map and vice versa (Caven, 2001).The purpose of their development was to combine previous map abilities into a single user-adjustable interface. This research is particularly relevant to mobile robot research because the requirements of successful navigation for robots is similar to that of humans, namely recognition and traversal of possible directions of travel (affordances) and recognition of obstacles.

Mobile robots can be used to perform tasks such as reconnaissance, surveillance and target acquisition, which demand higher situational awareness. Particularly, if such a task is complex, it is necessary to have an understanding of the spatial and functional properties of the environment where it operates. Due to inevitable restrictions in the field of view of the cameras on robots and in the quality of the visual displays feedback from remote robots, operators are often unable to maintain a level of situational awareness sufficient for safe and efficient task execution. According to Topp et al, the Human-Augmented Mapping (HAM) can help remote operators to acquire qualitative spatial knowledge based on a number of sensors on robots. Although with sensors equipped, tasks such as path planning and real-time mapping are hardly achieved by mobile robots. Augmented Reality can overlay critical digital information onto the real environment. However, sufficient amount of information is required to support AR to telerobotic exploration. The intelligent environment with sensors creates the ideal application context.

4 Intelligent Application Context

The intelligent environment, which is originally proposed to support human inhabitants, should be able to support robots as well. Hashimoto et al proposed an intelligent space where people and robots are inhabited together. Instead, the system presented in this paper separates robots from their remote operators. Robots are actually the remote representatives interacting with the intelligent environment. The proposed system can solve limitations of robots usability in an unknown environment.

The system architecture mainly consists of DIMM (Distribute Intelligent Monitor Manager), semi-autonomous robot and Augmented Reality visualization module (Fig 6).

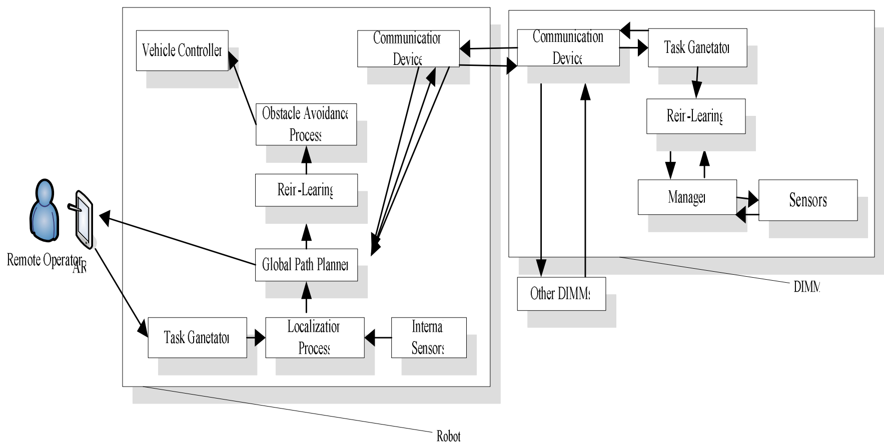


Fig. 6. Technical structure of the intelligent environment

Intelligent environment is divided into several sub-areas and each area is monitored by a DIMM node. DIMM has three main functions: detecting, managing, and communicating. The role of DIMM is to monitor events inside the environment through sensors. Based on different task requirements, the four main types of sensors equipped in each DIMM node are stereoscopic camera, ultrasonic range sensors, pyroelectric human monitor sensor and ambient temperature sensor.

Lee et al. (2005) have developed reinforcement learning as a computational approach of automating goal-directed learning and decision-making. The reinforcement learning module could help semi-autonomous robot to learn which action it should take in response to a state or rewards. The DIMM of the intelligent environment can monitor the position and velocity of moving objects, and then identifies a navigational map from a sensed situation. The mobile robot learns how to navigate in this situation based on the created map.

The mobile robot is semi-autonomous, which means it can be autonomous or telerobotic. When the robot is in the autonomous state, there is a global path generator using information from sensors that can make a new global path with reinforcement learning for the mobile robot while considering the state of the mobile robot, objects and other parameters in the environment. Otherwise, in the telerobotic state, those data will be sent back to AR visualization module for the remote operator to make decisions.

A request from the remote operator who is using AR visualization module to communicate with the robot is firstly translated into a task plan through a task generator. Following the requirements of the task, the communication device of mobile robot transmits these requirements to the DIMM that covers the current position of the robot. The DIMM will then synchronize the request to other DIMMs if necessary, after evaluating those requirements. Therefore, the robot can use sensors installed in all DIMMs areas which detect the dynamic position of the mobile robot, obstacles and objects.

A computational model was developed based on Merrick et al. (2007) to describe the characteristic reasoning process of DIMMs, robot and AR visualization module (the triangle interactions)(Fig 7).

The DIMM agent is a motivated reinforcement learning agent. The sensors for this agent, which are fixed into the environment, can detect the space and provide raw data to the sensation process. The motivation process produces a motivation value using sensed data from sensation process. The motivation value can be used by the learning process as an intrinsic reward signal to incorporate the previous parameters, actions and current reward into a rule that can is used to define what DIMMs should execute in the next step. Following this rule, the action process selects the option to trigger the effectors. For example, the DIMM can provide a propositional plan to the mobile robot.

The semi-autonomous robot agent can also be configured as a motivated reinforcement learning agent. The sensors for this agent include the sensors on the robot as well as for DIMMs. The basic process is similar to DIMMs agents. Before the rule is generated, the learning process needs to negotiate with the learning process of DIMM agent which provides a propositional plan.

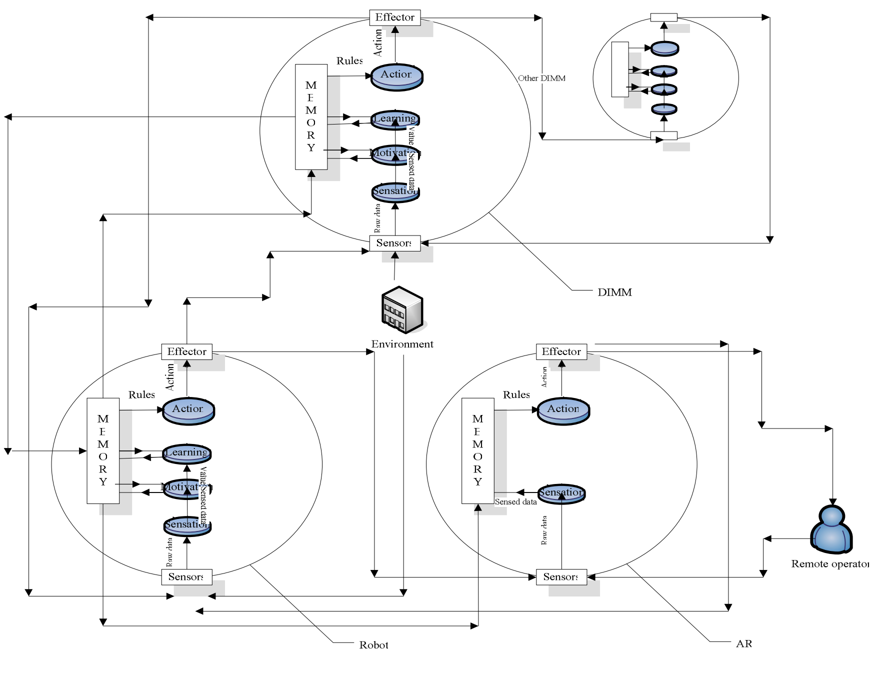


Fig. 7. The computational Agent model of the triangle relationship

Table 1. Sensors/effectors

Agent	Sensor/Effectors	Function
ROBOT	Rotary sensor	measure rotational movement.
	Acceleration sensor	measure both the static acceleration and robot
	Pyroelectric-human monitor sensors	incorporate pyroelectric infrared sensor to detect infrared energy radiation from human body
	Sharp IR distance measuring sensor	use infrared signal to measure object distance from 10-80 cm with analog output
	Communication effectors	communicate with DIMM and AR
	Control effectors	control the action of robot
DIMM	Ultrasonic range sensor	detect the range information
	Pyroelectric-human monitor sensors	incorporate pyroelectric infrared sensor to detect infrared energy radiation from human body
	Ambient temperature sensors	use temperature sensor to generate linear voltage signal according to the ambient air temperature.
	Stereopticon camera	capture SD image
	Communication effectors	communicate with DIMMs and robot
	Plan effectors	provide propositional plan to robot

The processes of the AR agent are facilitated by three modules (sensors, memory and effectors). Raw data is collected from the sensors on the robot and the DIMMs. The raw data is then transformed by the sensation process into sensed data which can trigger the effectors by the action process. The sensed data such as range, path and velocity can be used by the AR module to overlay digital information onto the video view of the remote operator. For example, the remote operator can easily control the robot to find the shortest way to a target based on a virtual navigational map created by the AR module.

The memory module of the robot, DIMMs and AR are under the control of a central database management system (DBMS) which periodically synchronize the scattered database to ensure that they all have consistent data through network. Triggering the triangle as shown in Fig. 16 can be any sensor associated with the three agents. The AR module can obtain the request from operators and initiate the circle. Alternatively, the sensors on the robot and DIMM can detect the changes of environment and therefore launch the circle. In the following discussions, it is assumed that the circle is triggered from the AR module upon the request from the remote operator: The sensation process of the AR module transforms the raw data from the operator into sensed data which is used by the action process to trigger effectors. The effectors on the AR module then send the request to the robot agent. Such transmission activates the sensors of the robot agent and then a motivation value can be obtained by the motivation process using the sensed data. The motivation value is a triggering of the learning process to make a rule to the action process. Actions following this rule could include moving the robot, requesting more information to the AR module, or sending requests to DIMM. In DIMM, the request is then processed to a motivation value to provide its leaning process an intrinsic reward. The learning process of the DIMM agent then makes a rule to its action process, which initiates the effectors (e.g., request more supports by sending requests to other DIMMs or provide a propositional plan to the robot). Such produced propositional plan can also be processed as intrinsic reward to the leaning process of the robot agent through the "memory", which takes account of previous parameters and actions to make a new rule. Alternatively, this propositional plan from DIMM can also be directly used by its own action process. The action process of DIMM then follows this rule to trigger its effectors which send data back to the AR module to be viewed by the remote operator. Table 1 lists the sensors/effectors and their corresponding functions for each agent in the system.

5 Summary

In order to significantly improve performance on a task with a teleoperated robot, it is important to improve the operator's situation awareness of the remote environment. Since the operator is not collocated in the same environment as the robot, the development of an operator's situation awareness must come through information visible on the user interface. Most of the current research in human-robot interaction focuses on how an operator could interact with a robot (i.e. using a PDA, gestures, the internet, a desktop computer, a head-mounted display) and what information could be useful to the operator (camera, range, map, proximity indicators, sensor status, waypoints, goals). However, the question of how the information should be presented to the operator has not been adequately addressed.

Conventional interfaces for teleoperating a remote robot do not adequately support the development of situation awareness because related information from the robot is usually presented in different parts of the display and the operator is responsible to mentally correlate the information. An interface that better supports the development of situation awareness would display related information in a single part of the display so that the operator can immediately observe how different sets of information are related to each other. Additionally, the effects of interfaces is typically validated by subjective evaluations or by showing that it fulfills a set of requirements or can be used to accomplish a particular task or set of tasks. The mixed reality interface correlate and present related information in a single part of the interface, thereby reducing the operator's cognitive workload required interpreting the information.

References

Boston, M.A., Gibson, J.J.: The ecological approach to visual perception. Houghton Mifflin (1979)

Woods, D.D., Tittle, J., Feil, M., Roesler, A.: Envisioning human-robot coordination in future operations. IEEE Transactions on Systems, Man, and Cybernetics, Part C 34(2), 210–218 (2004)

Casper, J., Murphy, R.R.: Human-robot interactions during the robotassisted urban search and rescue response at the world trade center. IEEE Transactions on Systems, Man, and Cybernetics, Part B 33(3), 367–385 (2003)

Lee, J.D., Caven, B., Haake, S., Brown, T.L.: Speech-based interaction with in-vehicle computers: The effect of speech-based email on driver's attention to the roadway. Human Factors 43, 631–640 (2002)

Burke, J.L., Murphy, R.R., Coovert, M.D., Riddle, D.L.: Moonlight in Miami: Field Study of Human-Robot Interaction in the Context of an Urban Search and Rescue Disaster Response Training Exercise (May 2004)

Drury, J.L.: Extending usability inspection evaluation techniques for synchronous collaborative computing applications. Sc.D. Dissertation, Computer Science Dept. University of Massachusetts Lowell (2001)

Drury, J.L., Yanco, H.A., Scholtz, J.: Using Competitions to Study Human-Robot Interaction in Urban Search and Rescue. ACM CHI Interactions, 39–41 (March/April 2005)

Wang, X., Kwok, N.: Inspecting Interior Environments Using Augmented Reality-based Autonomous Robots. In: Full paper accepted by the 24th International Symposium on Automation & Robotics in Construction (ISARC 2007), Kochi, Kerala, India September 19-21 (2007)

Wang, X., Dunston, P.S.: Mixed Reality — Enhanced Operator Interface for Teleoperation Systems in Unstructured Environment. In: CD Proceedings of the 10th Biennial ASCE Aerospace Division International Conference on Engineering, Construction and Operations in Challenging Environments (Earth and Space 2006), League City/Houston, Texas, March 5-8, 8 pages. American Society of Civil Engineers, ASCE (2006)

Alfano, P.L., Michel, G.F.: Restricting the field of view: perceptual and performance effects. Perceptual and Motor skills 70(1), 35–45 (1990)

Milgram, P., Kishino, F.: A taxonomy of mixed reality visual displays. IEICE Transactions on Information Systems E77-D(12), 1321–1329 (1994)

Zahorik, P., Jenison, R.L.: Presence as being-in-the-world. Presence 7(1), 78–89 (1998)

Murphy, R.R., Rogers, E.: Introduction to the special issue on human\–robot interaction. IEEE Trans. Syst., Man, Cybern. C, Appl. Rev. 34(2), 101–102

Santa Monica, C.A.: Design and evaluation for situation awareness enhancement. In: Endsley, M.R. (ed.) Proc. Human Factors Society 32nd Annual Meeting (1988)

Seattle, W.A., Lehikoinen, J., Suomela, R.: Perspective map. In: Proceeding of the Sixth International Symposium on Wearable Computers, ISWC (October 2002)

Sendai, J.P.N.: Ecological displays for robot interaction: A new perspective. In: Ricks, B.W., Nielsen, C.W., Goodrich, M.A. (eds.) Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (2004)

Hughes, S., Manojlovich, J., Lewis, M., Gennari, J.: Camera control and decoupled motion for teleoperation. In: Proceedings of the 2003 IEEE International Conference on Systems, Man, and Cybernetics, Washington, D.C. (2003)

Sheridan, T.: Musings on telepresence and virtual presence. Presence, Teleoperators, and Virtual Environments 1(1), 120–125 (1992)

Washington, D.C., Scholtz, J.: Human-robot interactions: Creating synergistic cyber forces. In: Proceedings from the 2002 NRL Workshop on Multi-Robot Systems (March 2002)

Yokohama, J.P.N., Milgram, P., Zhai, S., Drascic, D., Grodski, J.: Applications of augmented reality for human-robot communication. In: Proceedings of the International Conference on Intelligent Robots and Systems (IROS), pp. 1467–1472 (July 1993)

Computational Cognitive Modeling of Human-Robot Interaction Using a GOMS Methodology

D.B. Kaber[1], S.H. Kim[2], and X. Wang[1]

[1] Edward P. Fitts Department of Industrial and Systems Engineering, North Carolina State University, USA
[2] Department of Industrial and Manufacturing Systems Engineering, University of Michigan-Dearborn, USA

Abstract. The goal of this study was to use computational cognitive modeling to further understand human behavior and strategy in robotic rover control. To this end, GOMS (Goals, Operators, Methods, Selection Rules) Language models of rover control were constructed based on a task analysis and observations during human rover control trials. For the first model, we hypothesized control would be characterized by actions to prevent deviations from exact path following. The second model was developed based on an alternate hypothesis that operators commanded ballistic rover movements to approximate path direction. In manual trials, an operator was required to navigate a commercially available micro-rover along a defined path using a computer interface (providing remote environment information through a camera view) located in a room separate from the rover. The computational cognitive model was executed with a pseudo system interface (Java device) in real-time. Time-to-navigation completion and path tracking accuracy were recorded during the human and cognitive model trials with navigation circumstances being identical. Comparison of the GOMSL model outputs with human performance demonstrated the first model to be more precise than actual human control, but at the cost of time. The second model with the new navigation criteria appeared to be more plausible for representing operator behavior; however, model navigation times were consistently longer than the human. This was attributable to limitations of the modeling approach in representing human parallel processing and continuous control. Computational GOMS modeling approaches appear to have potential for describing interactive closed-loop rover control with continuous monitoring of feedback and corresponding control actions. Humans exhibit satisficing behavior in terms of rover direction and movement control versus minimizing errors from optimal navigation performance. Certain GOMSL modeling issues exist for applications to human-robot interaction and this research provides a first empirical insight.

Keywords: computational cognitive modeling, human-robot interaction, GOMSL, telerobotics, closed-loop control.

1 Introduction

Teleoperated (or remote-control) robots have been developed for a variety operating situations, including military and space applications, hazardous material

X. Wang (Ed.): Mixed Reality and Human-Robot Interaction, ISCA 47, pp. 53–75.
springerlink.com

handling, undersea welding operations, surgery, and search and rescue activities (e.g., see Micire (2007) for a contemporary example). Correspondingly, various human interfaces have been designed for operating telerobots (Yanco et al., 2007). Typical control stations include panel displays with robot camera views or map displays for operator perception of remote environments and hand-controls (keyboards, joysticks) for robot manipulation. Unfortunately, usability issues have been observed with such interfaces, including limited FOV (Field-of-View), camera viewpoint and orientation control (Chen et al., 2007), and incompatibility of display and control axes of reference, which can lead to poor human comprehension of telerobot states and constrained control behavior. For example, because of limited FOVs, operators may adopt a robot motion control strategy that focuses on minimizing deviations from a local path and preventing collisions with objects on the path versus using a ballistic control strategy and maintaining awareness of objects in a larger area (e.g., Krotkov et al., 1996).

In order to improve the effectiveness of human-robot interaction (HRI) for task performance, it is necessary to have a detailed understand how humans perceive and manipulate interfaces in controlling a robot. Developing models of human performance is one way to describe, interpret and predict internal behaviors (strategies) and external control actions. Drury et al. (2007) noted that while such modeling approaches have been established in the human-computer interaction (HCI) field, few studies have been conducted on how to adapt these modeling approaches to HRI. Drury et al. also offered that the GOMS (Goals, Operators, Methods, Selection Rules) modeling technique (developed by Card et al., 1983) is a popular, formalized HCI modeling method that may be well-suited for constructing HRI task models.

GOMS is a cognitive task modeling technique for representing user behaviors with interactive systems. It is used to quantify procedural knowledge necessary to operate a system. Computational forms of GOMS models can also be used to make quantitative predictions of user performance, such as task completion time. Early GOMS research focused on representing tasks with desktop computing applications. Gray et al. (1993) investigated telephone toll and assistance operator (TOA) tasks and workstation designs. Haunold and Kuhn (1994) used GOMS models to study human performance in map digitizing. These tasks and systems can be considered open-loop in nature. The visual display of information is predominately static and users can make decisions and act on information without requiring feedback on prior tasks. These characteristics have led to very similar GOMS analyses across operating conditions.

Related to this, other researchers have suggested that GOMS is not suitable for modeling behavior in real-time, closed-loop, dynamic and interactive environments. Agre and Chapman (1987) said such highly interactive tasks can only be modeled with reactive systems involving no planning whatsoever. More recently, Wei and Salvendy (2004) summarized that GOMS methods, including CPM (Critical path method)-GOMS and Natural GOMS language, are generally not capable of representing several cognitive attributes of human performance, including monitoring, communicating, attention and motivation, which are often part of closed-loop systems control. However, some empirical research has been conducted using GOMS to model closed-loop and interactive tasks in which operators required continuous feedback on system states as a basis for actions. John and Vera (1992) and Kieras and Meyer (1995) extended the scope of GOMS analyses through application to interactive video games and computer-based simulation of tactical military tasks. Like other domains where GOMS modeling has been successfully applied, expert behavior in these dynamic tasks initially appears to be highly complex and intense, but can ultimately be broken-down into routine sequences of actions applied at specific times. This research revealed procedure (planning)-based modeling methods can successfully describe behavior in highly interactive tasks. It is, however, important to note that the John and Vera's (1992) study was limited in terms of integration of the GOMS model with the task interface, specifically they did not use a complete task simulation. Their results on a 30s window of task run-time suggested that a GOMS model would be reactive and could simulate time-dependent processes.

The objective of the present work was to explore the use of computational GOMS models to develop an understanding of human cognitive strategy and behavior in remote rover control. In that many commercially available telerover technologies integrate basic alphanumeric or graphical user interfaces (GUIs; e.g., Evolution Robotics ER-1, iRobot Negotiator, MobileRobots Pioneer series) for closed-loop control tasks, we suspected that GOMS would be a plausible and effective way of representing cognitive processes (e.g., visual operations, long-term memory (LTM) and working memory (WM) use) and performance (e.g., task learning and execution times) in such HRI. There are many forms of GOMS models, ranging from the simple keystroke-level models (KLM) (Card et al., 1983) to complex CPM-GOMS models (Gray et al., 1993). There are flexible forms of GOMS coding, including NGOMSL (Natural GOMS Language) presented by Kieras (1997), and structured forms of coding like GOMSL (GOMS Language), also presented by Kieras (1998). Structured GOMSL coding allows for representation of user task methods and rules for decision making and action execution that can be input into a computer for simulation purposes using a complier (e.g. GLEAN (GOMS Language Evaluation and Analysis) tool). In this research, we evaluated the utility of GOMSL for modeling the highly interactive and closed-loop task of telerover control, and for analyzing the human internal control strategy (honing to a path versus ballistic motion control) by comparing model outputs with human data. Based on the prior teleoperated robot research, we initially hypothesized user rover control would be characterized by actions to prevent deviations from exact path following. For this research, we used a novel software tool developed by SoarTech (2005) called EGLEAN (Error-extended GLEAN) to code and debug the GOMSL models and facilitate direct model connections with a virtual device representing the rover control interface. This was an extension of John and Vera's (1992) research and we expected the GOMSL models to provide approximations of actual human performance in the interactive rover control task.

2 Method

2.1 Task

The specific task we modeled required human operators to control a micro-rover in tracking a predefined path in a laboratory setting. The complexity of the task was defined by the size and shape of the path, the number of turns (see Figure 1), as well as the speed and directional control mechanisms of the robot. We limited the area of the path to 240 × 240 cm, based on the size of the micro-rover and a time lag in remote control by an operator via a wireless network connection. The rover traveled at a fixed speed and there were no obstacles introduced into the path.

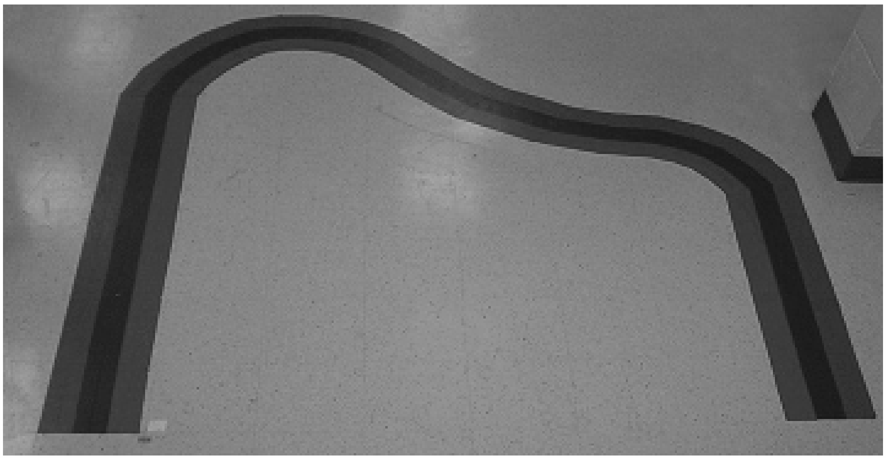


Fig. 1. Rover navigation path

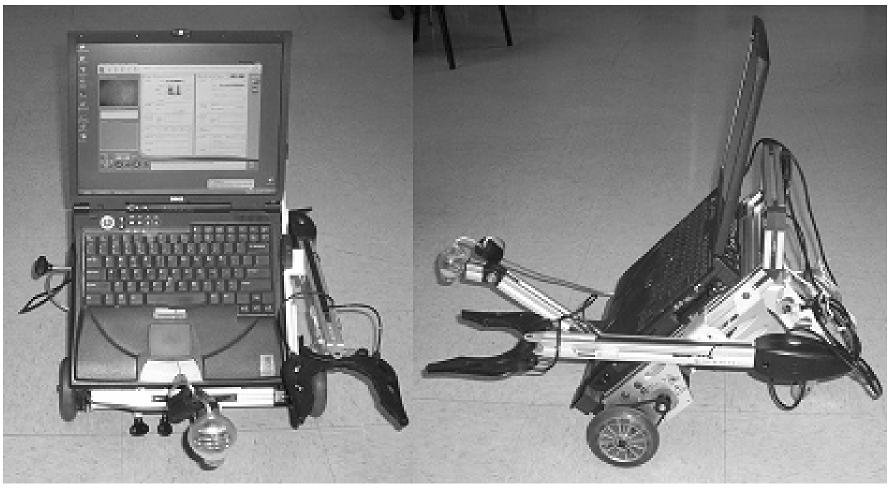


Fig. 2. ER1 robot

2.1 Robot Platform

The robot used in this study was an Evolution Robotics™ (ER) 1 personal robot system (see Figure 2). A laptop computer was mounted on the robot for wireless control and to allow for programming of the unit, including local control of sensors, an IP camera and IR sensors, and actuators (drive system and gripper).

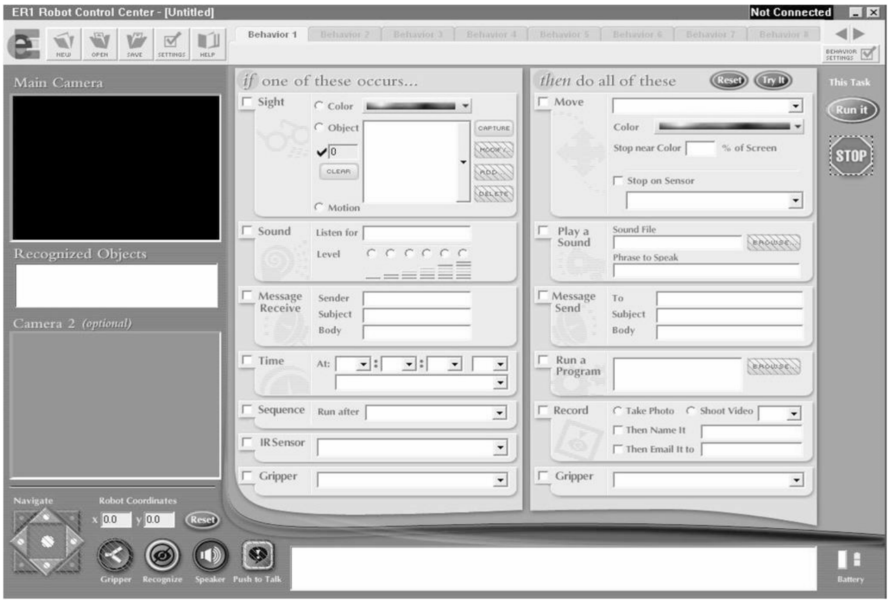


Fig. 3. RCC interface of the ER1

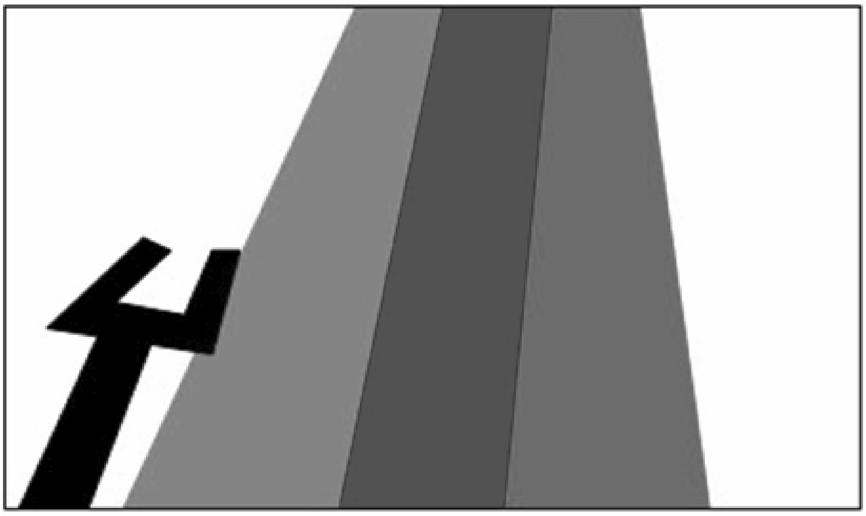


Fig. 4. Example graphic of camera view window

Remote control of the ER-1 also requires integration of a remote computer with Robot Control Center (RCC) software installed and connected to ER1 through a wireless network. The RCC software (see Figure 3) receives feedback from the robot sensors, including the camera system. Figure 4 shows a graphic of the camera view window in which human operators could see the path during tracking.

2.3 Cognitive Modeling Tools and Features

GLEAN provides the capability to compile and run cognitive models written in GOMSL (Kieras, 1998). It also provides the capability to conduct static analyses (task method counts, method execution profiles) on GOMSL model content as well as run-time analyses (WM load, execution times). GOMSL models include several basic components: operator goals, task methods, specific operations as part of methods and selection rule sets. By formulating selection rules according to individual preferences, GOMSL models are capable of representing the performance of individual operators (Card et al., 1983). GLEAN also supports use of a Decide operator to represent decision making processes within a specific method. Decide operators are for modeling simple decisions governing the sequence of actions according to a script in LTM.

2.3.1 Keeping Track of Time in GOMSL Models

GOMSL models describe strategies for a task by imitating human operator decision making processes. To evaluate the fitness of models, it is necessary to transform operation sequences to the time domain and make comparison with human performance times. GLEAN assigns a default execution time of 50 ms for every GOMSL statement. Extra time is added for various operators, for example, a mouse click (200ms). Kieras' (1998) manual for GOMSL presents the execution times for all currently available operators (or mathematical models for manually determining action/operator times in model evaluation). Kieras also provides some explanation for operator execution times based on historical psychology research (e.g., Card et al, 1983).

2.3.2 Psychological Constraints on GOMSL Models and Model Compilation

Psychological constrains are essential for constructing plausible cognitive models. Although theoretical constraints are difficult to represent or incorporate in computational methods, this serves as the foundation for GOMSL and other contemporary cognitive modeling approaches, such as EPIC (Executive Process Interactive Control) (e.g., Kieras & Meyer, 1995). As an example, values of task variables must either be retrieved from LTM or a Task definition, as part of the model, before related processing can occur. In addition, visual operators in GOMSL models assume that interface objects are within the area of foveal vision during processing. Objects outside foveal vision (e.g., in peripheral vision) can only be recognized after searching for them by shifting attention using specific model operators.

For complicated cognitive modeling applications, such as simulation of software interfaces and user states, GLEAN cannot be used in a "stand alone" manner. The complier must be integrated with additional C++ programs to implement notional representations of interfaces (identification of interface objects and their screen/display coordinates in code) for presentation to a GOMSL model. Furthermore, scenario/script files must be developed to populate the notional interfaces with data for model processing in run-time. Beyond these requirements, new operators cannot currently be defined for use in GLEAN (by end users), limiting its extension to some situations (e.g., mobile computing applications).

For these reasons, in the present study, we used EGLEAN (Error-extended GLEAN), which is a cross-platform system for human cognitive process simulation (Wood & Kieras, 2002), to facilitate GOMSL modeling and representation of the rover control interface (using Java). It works as a plug-in to the Eclipse IDE (Integrated Java Development Environment), a freeware programming tool. EGLEAN was developed based on the framework of GLEAN, thus it is implemented with all the psychological constraints and rules used in GLEAN. We also used GLEAN for specific model analysis purposes (e.g., task learning).

2.4 Experimental Approach

We compared human rover control in the path following task with cognitive model performance, when integrated with a Java device representation of the ER1 interface. (Each method of control is described below in detail.) We tested the operator in two trials and made observations on performance. We then conducted an unstructured post-trial interview and informal task analysis for GOMSL model development, including coding selection rules to represent the operator's internal criteria for controlling the rover and structuring model methods based on the task analysis. We implemented and tested two models including: (1) a stepwise control model; and (2) a ballistic motion control model in an attempt to account for human subject control behavior (Figure 5). The former model assumed the operator strategy to be one of exact path following. The latter approximated minimum task-time behavior with gross path following in comparison to the stepwise model. Since the rover navigation task scenario was the same in each trial, we only required the output of the computational cognitive model for one run (results did not vary from run to run).

2.4.1 Human Manual Control

A remote workstation was set-up in a room separate from that in which the rover was located and presented the RCC interface for the user. The operator could view the target path through the interface, as captured by the robot's camera, and control the rover to make turns or keep the vehicle moving forward. There were two different interface methods including (1) clicking a mouse pointer on appropriate controls or (2) using the arrow keys on a keyboard. These methods are highly representative of HCI tasks and motivated the applicability of the GOMSL model. During the test, the operator was instructed to complete the task as fast and as accurate as possible.

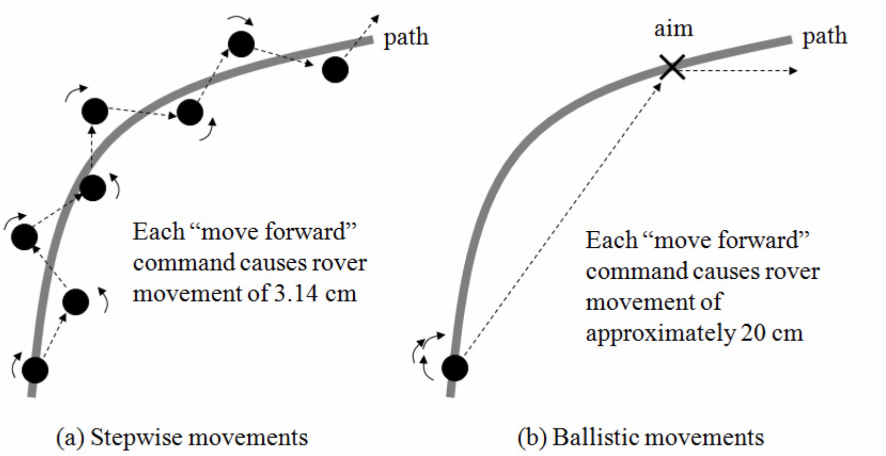


Fig. 5. Diagrams of stepwise (a) and ballistic (b) control movement hypotheses

In advance of the experiment, the operator was allowed to see the actual rover and the path in the lab space in order to develop a mental model for performance. The operator was also allowed to teleoperate the ER1 in several training trials until he was comfortable with the specific control methods and time lag. (This step was important because GOMSL models are intended to represent skilled/expert user performance). At the beginning of test trials, the ER1 was placed at the starting point on the target path and pointed in the correct direction. (Initial placement of the rover was also facilitated for the cognitive model trials.) The scope of the camera view was the same for the human performance and cognitive model trials. One difference in human manual control of the rover from the cognitive model was that the human accomplished the task in a parallel manner by looking at the camera view and, at the same time, executing robot control actions. The GOMSL model executed these behaviors sequentially. (This is one limitation of the simulated human cognitive processor in GLEAN, which we discuss later.) During both human performance trials and the cognitive model trials, navigation time and path tracking accuracy data were recorded.

2.4.2 GOMSL Model Control of Rover Navigation

2.4.2.1 Surrogate Java Device Interface

In the human performance trials, all visual feedback was provided via the rover camera. The visual processing required to numerically describe a real, dynamic scene to a computational cognitive model is quite complicated, and other work is just beginning to develop image processing components as part of cognitive architectures that may serve as bridges to pass data/information from dynamic environments to cognitive models (Ritter et al., in press). In order to quantify the information contained in the camera view in the ER1 interface (RCC software) for the GOMSL model, we developed a pseudo system interface using Java (see Figure 6). The cognitive model was linked to this device through the Eclipse IDE.

We measured the coordinates of points on the target path at every 5 cm and they were used to numerically define the path for the simulated rover device. All the coordinates were saved in a task scenario file, which was fed to the Java application to populate the pseudo interface with data. Once the simulated rover device moved within a certain range of the current target point (an event presented through the Java interface), the next point in the path was imported from the scenario file to simulate the process of the rover camera capturing the next real part of the path in the human performance trials. Once a new event was introduced, the information on the position of the rover, the current target point on the path, the distance of the rover from the target point and the directional angle relative to the current rover heading were updated.

By analyzing human control of the rover through the RCC interface, each button press to move forward resulted in an approximate displacement of 3cm and each button press to change direction resulted in an average angular rotation of the rover by 3.25 degrees. Consequently, these values were associated with model control actions for forward, left and right movement at the pseudo system interface.

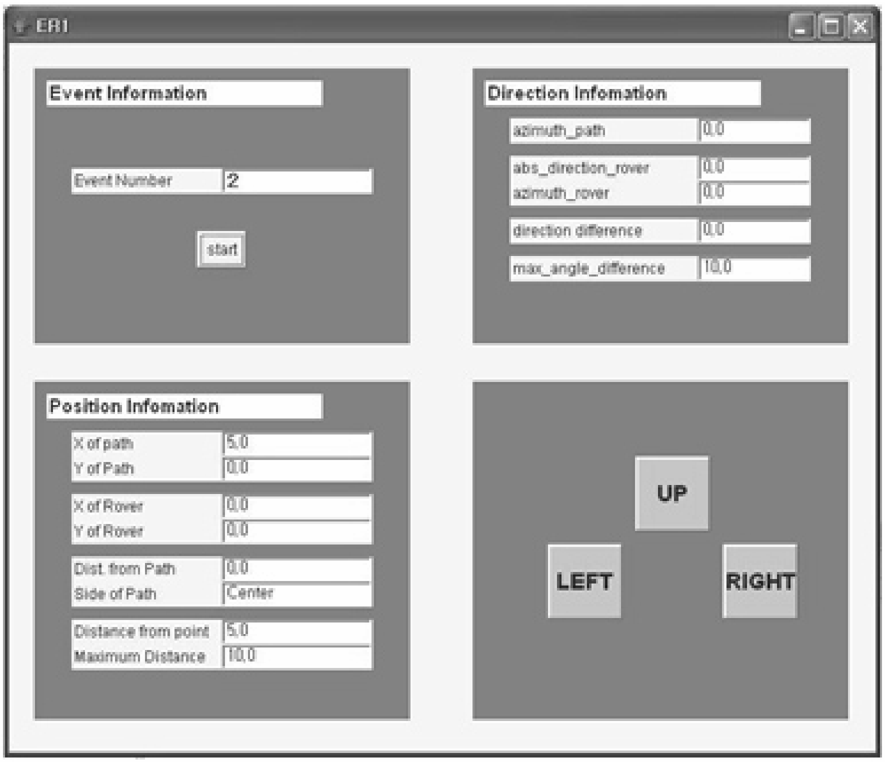


Fig. 6. Pseudo RCC interface developed in Java

2.4.2.2 Stepwise GOMSL Model Coding

Our stepwise GOMSL model included “Navigate ER1”, as the overall goal (or the top level method). Two criteria were used to define tolerances for acceptable accuracy in path tracking, including: (1) the distance of the rover from the target point on the path at the current position; and (2) the angular deviation of the rover direction from the path,. If the rover distance from the point on path or the deviation in direction exceeded the criterion values specified in the GOMSL model, the model output appropriate behaviors to move the rover forward or correct the direction. These criteria were based on the displacement deviation of the rover from a point on the path ahead that the human operator typically considered acceptable and the degrees of angular deviation from the path the operator considered critical in order to control the rover accurately. Accordingly, two second-level methods were coded in the model, including “Evaluate distance” and “Evaluate angle” (see the Appendix for a portion of our model code). For each of these methods, a selection rule set was defined so the model could decide whether the ER1 should turn left or right or move forward.

With respect to the decision criteria used in our model, it is usually neither practical nor reliable to collect extensive human behavioral data to quantify specific judgment criteria used in control tasks. Task experts often have difficulty in explicitly describing task goals, strategies and mental processes through knowledge elicitation (Gordon & Gill, 1994). For example, it would be very difficult for an operator to answer questions such as, “how much distance do you consider acceptable between the rover and the next point on the path visible through the camera view?”, “how many degrees of angular deviation of the direction of the rover from the target path do you consider critical?”; or “which turns should be made to maintain accurate rover tracking?” In many cases, the analyst must make judgment calls on criteria reflected in the selection rules of a GOMS model. Thus, for this model, we hypothesized that the operator used tight path deviation criteria for performance.

We tested the accuracy of the model and made modifications regarding some implementation issues. We determined that it was necessary to separate the two decision criteria (distance and angular direction) used for rover control across the GOMSL model and Java device. Since the Java device allowed us to set a criterion distance for updating event information in the pseudo interface, the decision criterion for aiming the rover at a point on the path could be handled by the Java device as well (see Figure 7 for simplified flow diagram). Accordingly, only one selection rule set for rover direction control was coded in a revised stepwise control model. This separation of the decision criteria across the applications allowed the GOMSL model size to be reduced and to run quicker, relative to human performance, as a result of reduced variable handling.

In order to assess the accuracy of the criteria coded in the Java device and the GOMSL model, we conducted a sensitivity analysis using various values of the decision criteria and compared model outputs with the tracks for human manual control. Based on our observations of the operator performance, we tested models including 3, 5, and 10cm as acceptable distance criteria and 3, 5, and 10 degrees of rover direction deviation as critical values in the model selection rule set. Finally, the stepwise GOMSL model was further refined by looking for redundancy in the steps coded as part of the methods. Methods with similar steps were decomposed and lower-level basic methods were developed.

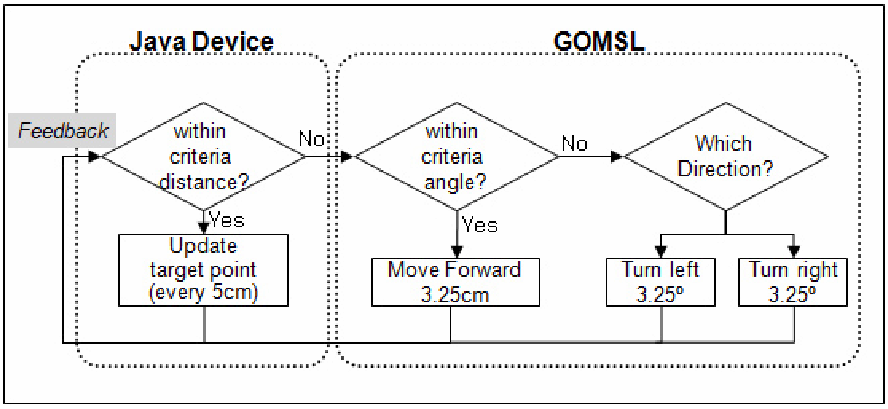


Fig. 7. Simplified action flow diagram for model control of the rover interface

2.4.2.3 Ballistic GOMSL Model Coding

We compared the stepwise control model with the ballistic strategy model for describing actual user behavior. The ballistic GOMSL model was created based on our observations of behaviors of three additional human subjects and modifications of several of our assumptions on task performance. We re-ran our experiment and conducted post-trial interviews with the additional subjects. As in the first experiment, the subjects were also asked to complete the task as fast and accurate as possible. In general, they tended to control the rover in a "ballistic" manner. They typically set the direction of the rover along the path by looking at a point in the camera image, approximately 20cm ahead of the rover's current position. They then pressed an interface button to turn left or right to direct the rover at the point. Similarly, once an operator aimed the rover at the path point in the camera view, they made a "bee-line" for the point until the rover had traveled, on average 20cm from its previous position.

In the ballistic model, we also added a decision condition to determine the degree of rover rotation from a particular position, required to align the vehicle with the target point on the path. Once the required direction and number of button presses for changing the direction was determined, the method of "turn left" or "turn right" in the GOMSL model fired multiple times, clicking the interface buttons as a human operator would. We also modified the GOMSL model method, "move forward", to advance the rover approximately 20cm with each press of the "up" arrow in the interface by the cognitive model. (A portion of the ballistic model code is included in the Appendix.) In addition to this, the path points in the scenario file (to be read by the Java device) were re-coded for every 20cm. In the ballistic model, it was not necessary to consider the criterion for evaluating the angular deviation of the rover from the path. The model was coded to change the direction of the rover and aim as close as possible at the target point after every forward movement command. However, the criterion for deciding the acceptable distance between the current rover position and the target points was retained and set to 20, 30 and 50cm in different versions of the model. Since the path points in the scenario file were written for every 20cm of path, it was no possible to set the distance criterion under 20cm.

With these modifications in the GOMSL model, we expected a substantial reduction in the predicted task completion time relative to the stepwise control model. The stepwise model represented an operator looking at the RCC interface with every control action (move forward, turn left, or turn right). If the perceptual requirement for tracking a rover destination was reduced based on operator control behavior, the total task time should also decrease.

In the cognitive model trials, the Java device and EGLEAN generated output data including the task completion time (quantified as minutes from the rover start position to the end of the marked path), the steps in the task sequence, and rover position points. We used the latter output to determine model path-tracking error, calculated as the mean accumulated distance deviation of the actual rover path from the target path. We compared these data with the same measures on human performance averaged across the multiple test trials and compared the fit of the two different models.

3 Results

3.1 Navigation Time

Table 1 presents the mean rover navigation time for the human operator and the stepwise GOMSL model with the various decision criteria. It is important to note that the GOMSL model output times were transformed to account for redundant keypresses in controlling rover motion not representative of actual operator behavior. For each point in the target path at which the cognitive model "decided" to move the rover forward (~3 cm), it executed a keypress for such action. However, when the rover was traveling in a straight line, the human operator merely pressed the appropriate control key once and "held it down" until the rover achieved an intermediate positional objective, mentally defined by the operator. This was also the case when the operator controlled the rover to execute a semi-continuous left or right turn as part of the path. Although the GOMSL modeling technique supports a "hold-down" operator to represent holding an interface button for some time, it was not possible to use this operator in our integrated modeling work with the Java device and EGLEAN and achieve real-time system control. We accounted for this through the systematic post-trial data filtering. The simple transformation of the navigation times led to more accurate model approximation of actual operator performance times.

Table 1. Rover navigation trial times for stepwise control model

	Rover Control Mode			
Time (m)	Manual	Stepwise GOMSL Model		
	mean	d=3, a=3	d=5, a=5	d=10, a=10
	1.65	19.95	16.83	15.81

The stepwise control model took substantially longer to complete the task than the human operator; however, there was a reduction in navigation time when the deviation criteria for distance from the target point and rover direction were relaxed. In general, the relaxed criterion model appeared to improve, but was still much slower than actual manual control. Table 2 presents the mean rover navigation time for the human operators (averaged across two trials × three participants) and the ballistic GOMSL models with various distance criteria. While the navigation time for our stepwise model was 10 times greater than the navigation time for the human operator, the navigation time using the ballistic model was reduced to less than 2 times that of the human. There was also a reduction in navigation time when the distance criterion in the ballistic model was relaxed.

Table 2. Rover navigation trial times for ballistic control model

	Rover Control Mode			
Time (m)	Manual	Ballistic GOMSL Model		
	mean	d=20	d=30	d=50
	1.73	4.56	3.95	3.44

3.2 Rover Path Tracking Deviations

The mean values of accumulated distance deviation of the actual rover tracks from the target path for human performance trials are presented in Table 3 along with the virtual path tracking accuracy for the stepwise control model runs. The values in the table reveal that the cognitive model was far more accurate than the human operator, but there was also a slight increase in rover distance deviation from the path when the distance and direction criteria were relaxed.

Table 3. Rover path tacking accuracy for stepwise control model

	Rover Control Mode			
Deviation (cm) for entire path	Manual	Stepwise GOMSL Model		
	mean	d=3, a=3	d=5, a=5	d=10, a=10
	1652.85	237.06	254.92	303.88

The patterns of rover path tracking for the human and stepwise control model trials are presented in Figures 8 and 9. The stepwise GOMSL models produced highly precise performance, even when the most relaxed directional error criterion was used. In general, the smaller control criterion, the more precise the model tracking was.

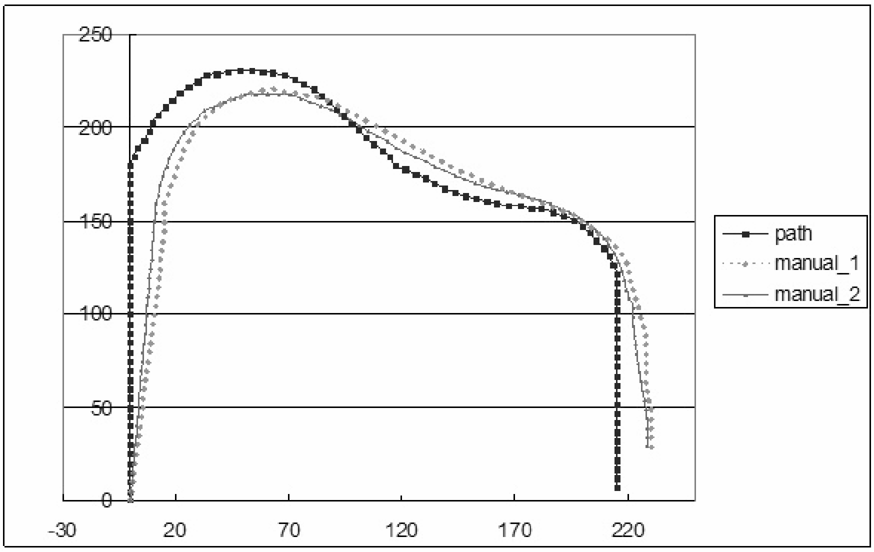


Fig. 8. Tracks of manual mode trials

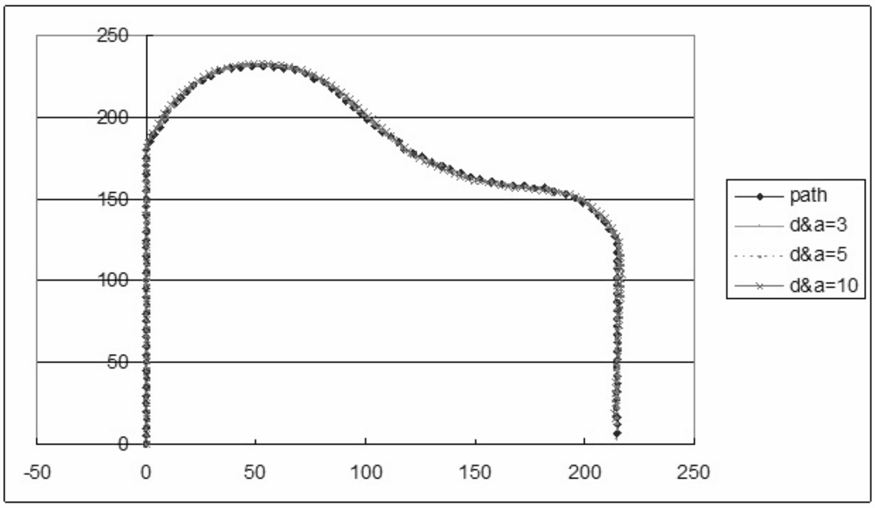


Fig. 9. Tracks for stepwise GOMSL model runs

From these data, we inferred that a speed-accuracy tradeoff existed in model performance and that the two criteria for rover control could be relaxed even further to allow for more accurate prediction of human performance (as in the ballistic control model). However, the speed results suggested that the law of diminishing returns might apply to reductions in task time for the cognitive model with further loosening of the tracking accuracy criteria. (We ran models with d=20, a=20 and d=30, a=30 and the performance times were 15.51m and 14.34m, respectively.)

Opposite to the stepwise model results, the accumulated distance deviations for virtual tracking accuracy for the ballistic GOSML model approximated human performance much better. Table 4 reveals the deviation for the cognitive model, given a 50cm distance criterion, was very close to the path deviations for the humans. Figures 10 and 11 present the tracks for the ballistic GOMSL model runs with various distance criteria and the human operator runs. From the figures, it can be seen that the pattern of output for the model with 50cm distance criteria was very similar to the pattern of the tracks for the human manual trials.

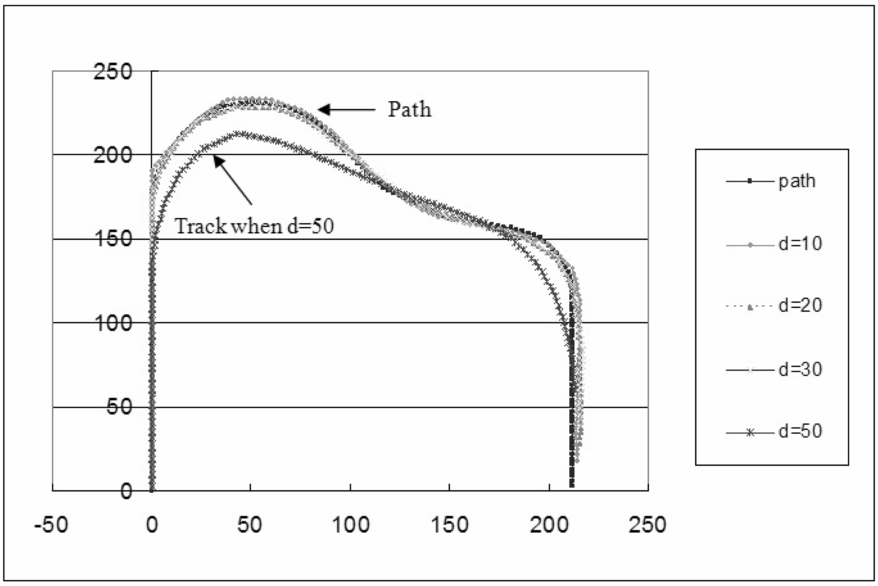


Fig. 10. Tracks for ballistic GOMSL model runs

In general, the results indicated that the modified set of assumptions on internal decision criteria for ballistic control of the rover generated output much closer to human performance. The model predicted human control behavior with approximately 1% difference in tracking accuracy; however, the speed of the model was 1.7 min slower than the humans (attributable to the serial processing limitation of the EGLEAN compiler). Based on the improvement in model performance and further comparison with human performance, it can be inferred from these results that human operators predict in advance the pattern of the path of the rover during control. This may depend on the degree of “look ahead” they have from a camera view.

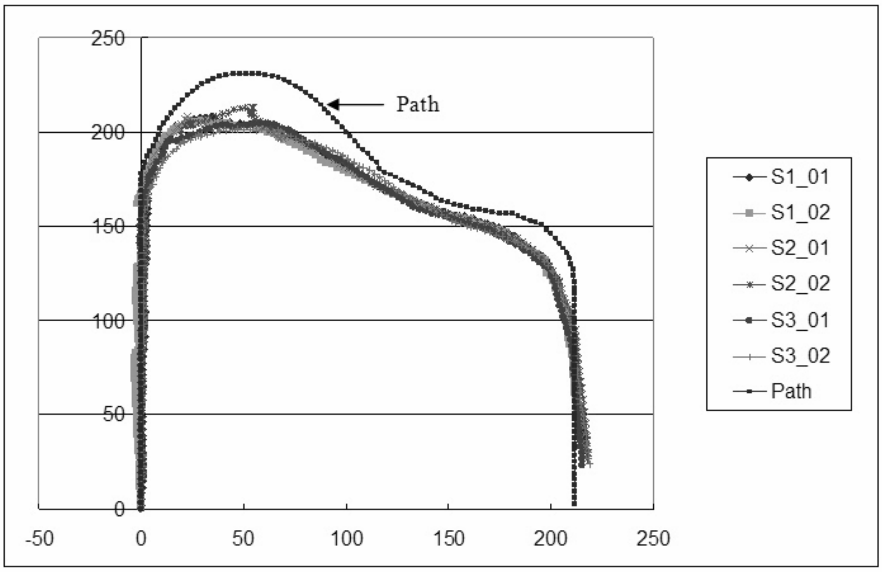


Fig. 11. Tracks of second manual mode trial

Table 4. Rover path tacking accuracy for ballistic control model

	Rover Control Mode			
Deviation (cm) for entire path	Manual	Ballistic GOMSL Model		
	mean	d=20	d=30	d=50
	1562	365	470	1579

3.3 Discussion

Although the stepwise model succeeded in driving the rover through a target path with the pseudo interface in real-time, the results revealed substantially different speed and accuracies for the human versus the model. Such differences are beyond the historical cognitive model validity criterion recommended by John and Newell (1989) of 20% deviation between model output and human performance. However, the ballistic model showed significant improvement regarding the performance time. Related to this, Table 5 presents the method execution profiles for the ballistic GOMSL models we tested, including the percentage of total task time during which each goal or high-level method was active. Based on the simplicity of the task and the ER1 control interface, the majority of methods included in the model were low level (“Move forward”, “Turn left” and “Turn right” methods); that is, used to describe interface behaviors.

The Table reveals rover motion control methods were active roughly 30% of the task time; thus, perceptual and memory operators, including looking for task objects and storing information, account for the remainder of the execution time (70%). On this basis, it can be inferred that the majority of task time in rover navigation is spent on recognizing information in the remote environment versus performing interface actions. It can also be seen from Table 5 that the “Adjust angle” method was most time-consuming, at 30% of total processing time. This indicates that operators may spend more time on directional control versus exact positioning of a rover. Because of the configuration of the target path, the “Move forward” methods accounted for much more time than the “Turn left” or “Turn right” methods. Since there were two right turns and one left turn in the target path, the “Turn right” methods fired more than “Turn left” methods. It can also be seen from the Table that as the criterion for rover deviation from the path was relaxed, both the “Turn left” and “Turn right” methods were executed less.

Table 5. Method execution profiles for ballistic GOMSL models

Goal/Method	Models (with different decision criteria)		
	d=20	d=30	d=50
	Percent Activation (%)		
“Navigate ER2”	**100**	**100**	**100**
“Adjust angle”	**29.39**	**30.44**	**30.70**
“Move forward”	**17.55**	**19.19**	**20.55**
“Turn left (n)”	**1.08**	**1.06**	**0.25**
n=3.25	0.26	0.25	.
n=6.5	0.48	0.47	0.25
n=10	0.35	0.34	.
n=20	.	.	.
n=40	.	.	.
“Turn right (n)”	**7.49**	**7.35**	**6.92**
n=3.25	0.91	0.89	0.69
n=6.5	1.19	1.16	1.52
n=10	1.38	1.35	2.94
n=20	4.02	3.94	0.71
n=40	.	.	1.06

4 General Discussion

From the experiment and modeling results, we found that humans tend to exhibit “ballistic” control behavior, versus “stepwise” control, in rover navigation by aiming at a target point within a visible area about a defined path through a camera view. Even though the overall rover path-tracking task can be categorized as a closed-loop compensatory tracking task, operators may consider a single rover movement to be an open-loop control task consisting of aiming, redirecting and moving forward. On this basis, it can be inferred that human operators generate an “optimal” internal strategy to balance the speed-accuracy tradeoff.

Although the path tracking accuracy predicted by the ballistic model was very accurate, the speed results for the model still deviated from the human performance results by more than the 20% (the model validity criterion recommended by John and Newell (1989)). There are several potential reasons for this. First, the human operators performed all control actions in a parallel manner. The operator was able to look at the feedback from the camera and control the rover at the same time. A know limitation of GOMSL modeling is that it represents sequential performance of (perceptual and motor) methods. Other cognitive modeling approaches have been developed to address this issue, including CPM-GOMS (see John, 2003), which can simulate parallel information processes (vision, manual control, etc.). In EGLEAN, the visual perception operators and rover control operators were executed sequentially. Consequently, the model accomplished the task in a serial manner instead of applying continual control as the human did. This difference contributed to the long execution times for the GOMSL model.

Second, at the keystroke-level of the GOMSL model, the existing operators may not be sufficient to represent actual HRI in real-time tasks. Given the similarity in interface technologies for HCI (computing) tasks and HRI applications, this type of insight needs to be empirically-derived and not based on analytical assessments of the potential of GOMS for modeling HRI tasks, as Drury et al. (2007) have presented. Unlike the human operator’s keypress and hold behavior, the move forward control process for the GOMSL model was achieved by repeatedly clicking on the appropriate keyboard key. This was also the case for the “Turn left” and “Turn right” control processes, and for this reason we transformed the cognitive model performance times. In order to solve this problem in the future, additional and more robust interactive control operators, and real-time implementation techniques, need to be integrated in GOMSL and GLEAN for better modeling of HRI tasks. In addition to this, as Wagner et al. (2006) indicated, some primitive GOMS operator execution times may not be correct for this particular type of task.

Related to the previous model execution issue, there also appears to be a tradeoff in using a structured programming format in GOMSL modeling of interactive tasks with execution time prediction. Like other classical programming languages, a structural programming style is recommended for creating GOMSL models. This means that low-level operators, which are used to achieve higher-level methods or goals, should be classified as lower-level unit tasks. However, as we mentioned above, such structural programming leads to additional statements in the model, which increases the execution time (50 ms/statement). In situations where a higher-level method is called many times, the extra time associated with activation of lower-level methods accumulates and makes a significant difference in the final model execution time. In other words, there appears to be a tradeoff between the clarity of the program structure and the efficiency of the model.

Beyond the above issues, since all the feedback in the rover control task was provided to the human through the on-board camera view and the GUI, understanding how the human operator extracts useful information from the camera images is critical for construction of a plausible GOMSL model. Unfortunately, we only have a partial understanding of how humans perceive and process graphical information in complex recognition tasks (e.g., Darken & Sibert, 1996). There has been some cognitive modeling-related research towards identifying the task environment properties that are extracted and used by rover operators for manual control of navigation (St. Amant et al., 2005); however, it remains unclear what information from the environment is transformed to input information. Related to this, in the first GOMSL model we developed for our task, two criteria (rover distance from the target path and directional error) were used while only one criterion was actually required in the second model. At this point, there is no theoretical way of saying which model more closely represents the actual cognitive process of the user. Therefore, we must make comparison of the output of GOMSL models with human performance in order to gain insight into actual cognitive behavior.

Finally, unlike humans, GOMSL models simulate error-free performance throughout a task. The output for a model will always be the same given the same set of decision criteria in selection rule sets and the same scenario/input file. In reality, human operators cannot generate exactly the same performance in a task, like rover navigation control. For human operators, it is likely that the critical values used for tracking judgments are not precise and constant. The values may change under different situations and, consequently, the content of any GOMSL model for performance prediction or interface evaluation would need to change for accuracy. Beyond this, it may be possible to represent error-like behaviors in GOMS model output occurring with a certain probability (Schrepp & Fischer, 2006). By introducing variability in method operator time estimates (as Card et al. (1983) presented), it may be possible to produce stochastic GOMS model output closer to human operator behavior.

5 Conclusion

The objective of this research was to demonstrate the use of computational GOMSL modeling to understand human operational strategy in telerover path tracking control, by making comparisons of actual human performance and cognitive model output. No prior work, including Drury et al. (2007) or Wagner et al. (2006), has taken a computational approach to GOMSL modeling of HRI tasks or applied GOMSL models directly to real task interfaces using EGLEAN and Java devices. These are critical steps in cognitive model systems development and integration that led to the inference that humans tend to use “ballistic” control or rover motion within a camera view area. Although our final results revealed different speeds for the human versus the GOMSL model, the model succeeded in driving the rover through a complex target path and the pattern of control and degree of accuracy mimicked that of actual operators.

It appeared from our results that there is some utility to developing cognitive models to represent behavior in HRI tasks, and the possibility of applying them for real-time systems control versus using traditional automated control algorithms. The importance of this work is that this has not been previously empirically demonstrated. A cognitive model provides flexibility and robustness in modeling human behavior along multiple performance dimensions and can allow for critical insight into operator internal strategies and decision criteria. A valid cognitive model, as a basis for automated rover control, might outperform a basic autonomous control algorithm using an integrated vision system with existing robot control software providing for discrete movement commanding. Future research should make empirical comparison of such rover control modes.

In general, this study supported the contention that computational GOMSL methods can be used to represent highly interactive and closed-loop control tasks. However, our modeling experience revealed certain limitations of GOMSL relevant to the HRI domain, which have not previously been identified and may account for the majority of differences in human performance and model output. Specifically, existing low-level GOMSL operators may be insufficient for describing complex interface actions and it is desirable to support parallel processing of unit tasks in GOMSL to improve model results, relative to human behavior. These limitations of GOMSL for predicting task time should be addressed in future modeling research.

Appendix: GOMSL Codes

A. Stepwise Gomsl Model

Define_model: "ER1 auto-navigation"
Starting_goal is Navigate ER1.

//Top level method
Method_for_goal: Navigate ER1
Step 1. Look_for_object_whose Label is Max_distance and_store_under
<max_dis>.
Step 2. Look_for_object_whose Label is event and_store_under <event>.
.
.
.
Step 7. Accomplish_goal: Adjust angle.
Step 8. Goto 2.

Selection_rules_for_goal: Adjust angle
If <angle> is_greater_than "-10", and <angle> is_less_than "10", Then
Accomplish_goal: Move forward.
.
.
.

Return_with_goal_accomplished.
Method_for_goal: Move forward
Step 1. Point_to Up_arrow_key.
Step 2. Click Up_arrow_key.
Step 3. Return_with_goal_accomplished.

B. Ballistic Gomsl Model

Define_model: "ER2 auto-navigation"
Starting_goal is Navigate ER2.

//Top level method
Method_for_goal: Navigate ER2
Step 1. Look_for_object_whose Label is Distance_from_point and_store_under <distance_from_pt>.
..
Step 3. Look_for_object_whose Label is Angle_difference and_store_under <angle_difference>.
.
.
.
Step 6. Goto 1.

Selection_rules_for_goal: Adjust Angle using <angle>
If <angle> is_greater_than "-3.25", and <angle> is_less_than "3.25", Then Accomplish_goal: Move forward.
.
.
.
If <angle> is_less_than_or_equal_to "-3.25", Then Accomplish_goal: Turn_right 3.25.
Return_with_goal_accomplished.

Method_for_goal: Move forward
Step 1. Click Up_arrow_key ; Click Up_arrow_key ; Click Up_arrow_key ; Click Up_arrow_key ; Click Up_arrow_key; Click Up_arrow_key.
Step 2. Return_with_goal_accomplished.

Acknowledgement

We thank SoarTech Corporation, including Scott Wood, Keith Knundsen and Toby Tripp, for support in our use of EGLEAN. We thank David Kieras for his support in updating GLEAN3 to GLEAN4 for our use, and that of others. We also thank Thomas Horton and Lloyd Williams for their assistance in developing the

Java device representation of the RCC software. David Kaber and Sang-Hwan Kim's work on this project was supported by a grant from NASA Langley Research Center (no. NNL05AA20G). Kara Latorella was the technical monitor. The opinions expressed in this chapter are those of the authors and do not necessarily reflect the views of NASA.

References

Agre, P.E., Chapman, D.P.: Pengi: An implementation of a theory of activity. In: Proceedings of the Sixth National Conference on Artificial Intelligence, AAAI 1987, pp. 268–272. Morgan Kaufmann, Milan (1987)

Card, S.K., Moran, T.P., Newell, A.: The psychology of human-computer interaction. Erlbaum, Hillsdale (1983)

Chen, J.Y.C., Haas, E.C., Barns, M.J.: Human performance issues and user interface design for teleoperated robots. IEEE Transactions on System, Man, and Cybernetics-Part C: Applications and Review 37(6), 1231–1245 (2007)

Darken, R.P., Sibert, J.L.: Wayfinding strategies and behaviors in large virtual worlds. In: Proceedings of ACM SIGCHI 1996, pp. 142–149. ACM, New York (1996)

Drury, J.L., Scholtz, J., Kieras, D.E.: Adapting GOMS to model human-robot interaction. In: Proceedings of the ACM/IEEE international conference on Human-robot interaction 2007, Arlington, VA, pp. 41–48 (2007)

Gordon, S.E., Gill, R.T.: Cognitive task analysis. In: Zsambok, C.E., Klein, G.A. (eds.) Naturalistic Decision Making, pp. 131–140. Lawrence Erlbaum & Assoc., Mahwah (1994)

Gray, W.D., John, B.E., Atwood, M.E.: Project Ernestine: A validation of GOMS for prediction and explanation of real-world task performance. Human-Computer Interaction 8(3), 209–237 (1993)

Haunold, P., Kuhn, W.: A keystroke level analysis of manual map digitizing. In: Campari, I., Frank, A.U. (eds.) COSIT 1993. LNCS, vol. 716, pp. 406–420. Springer, Heidelberg (1993)

John, B.E.: Information processing and skilled behavior. In: Carroll, J. (ed.) HCI models, theories and frameworks, pp. 55–101. Morgan & Kaufmann, Amsterdam (2003)

John, B.E., Kieras, D.E.: Using GOMS for user interface design and evaluation: Which technique? ACM Transactions on Computer-Human Interaction 3(4), 287–319 (1996)

John, B.E., Newell, A.: Cumulating the science of HCI: From S-R compatibility to transcription typing. In: Proceedings of CHI 1989, pp. 109–114. ACM, New York (1989)

John, B.E., Vera, A.D.: A GOMS analysis of a graphic, machine-paced highly interactive task. In: Proceedings of the CHI 1992 Conference on Human Factors in Computing Systems, pp. 626–633. Addison-Wesley, New York (1992)

Kieras, D.E.: A guide to GOMS model usability evaluation using NGOMSL. In: Helander, M., Landauer, T.K., Prabhu, P. (eds.) Handbook of human-computer interaction, pp. 733–766. Elsevier Science, Amsterdam (1997)

Kieras, D.E.: A guide to GOMS model usability evaluation using GOMSL and GLEAN3 (Tech. Rep. No. 38, TR-98/ARPA-2). University of Michigan, Electrical Engineering and Computer Science Department, Ann Arbor, MI (1998)

Kieras, D.E., Meyer, D.E.: Predicting human performance in dual-task tracking and decision making with computational models using the EPIC architecture. In: Proceedings of the 1995 International Symposium on Command and Control Research and Technology, National Defense University, Washington, D.C (1995)

Krotkov, E., Simmons, R., Cozman, F., Koenig, S.: Safeguarded teleoperation for lunar rovers: From human factors to field trials. In: Proc. IEEE Planetary Rover Technology and Systems Workshop (1996)

Micire, M.: Evolution and field performance of a rescue robot. Journal of Field Robotics 25(1), 17–30 (2007)

Schrepp, M., Fischer, P.: A GOMS model for keyboard navigation in web pages and web applications. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) ICCHP 2006. LNCS, vol. 4061, pp. 287–294. Springer, Heidelberg (2006)

Soar Technology: EGLEAN Science and Technology Report: The first six months. AUTHOR, Ann Arbor, MI (2005)

St. Amant, R., Riedl, M.O., Ritter, F.E., Reifers, A.: Image processing in cognitive models with SegMan. In: Proceedings of the HCI International (2005)

Wagner, A.R., Endo, Y.E., Ulam, P., Arkin, C.: Multi-robot user interface modeling, . *In: Gini, M., Voyles, R. (eds.) Proceedings of the DARS 7, pp. 237–348 (2006)*

Wei, J., Salvendy, G.: The cognitive task analysis methods for job and task design: review and reappraisal. Behavior & Information Technology 23(4), 273–299 (2004)

Wood, S.D., Kieras, D.E.: Modeling human error for experimentation, training, and error-tolerant design. In: Proceedings of The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC), U.S. Army, Orlando (2002)

Yanco, H.A., Keyes, B., Drury, J.L., Nielsen, C.W., Few, D.A., Bruemmer, D.J.: Evolving interface design for robot search tasks. Journal of Field Robotics 24(8-9), 779–799 (2007)

Mental Transformations in Human-Robot Interaction

B.P. DeJong[1], J.E. Colgate[2], and M.A. Peshkin[2]

[1] Central Michigan University, USA
[2] Northwestern University, USA

Abstract. Human-robot interfaces can be challenging and tiresome because of misalignments in the control and view relationships. The human user must mentally transform (e.g., rotate or translate) desired robot actions to required inputs at the interface. These mental transformations can increase task difficulty and decrease task performance. This chapter discusses how to improve task performance by decreasing the mental transformations in a human-robot interface. It presents a mathematical framework, reviews relevant background, analyzes both single and multiple camera-display interfaces, and presents the implementation of a mentally efficient interface.

Keywords: Mental transformation, control rotation, control translation, view rotation, teleoperation.

1 Introduction

In the summer of 1997, Argonne National Laboratory spent 2000 man-hours and $1.38 million dismantling their recently decommissioned nuclear reactor (Department of Energy, 1998). Rather than place humans in the radioactive environment, Argonne used a remotely controlled robotic system called the Dual Arm Work Platform (DAWP), consisting of two six-degree-of-freedom robotic arms and several tilt/pan/zoom cameras (Figure 1). Human operators sat at a console with several fixed video monitors and controlled the robots via two passive manipulanda[1] (Noakes et al., 2002).

The use of teleoperation[2] was cost-effective, but Argonne personnel noted several problems. First, training the operators was time-consuming and expensive; only 60% of the tested operators were skilled enough to complete tasks. Second, operators spent nearly 90% of the their time prepping rather than performing tasks. Finally, the teleoperation was mentally tiring, especially when performing complex tasks that required switching between multiple camera views (DeJong et al., 2004b).

[1] Manipulandum is a general term for the device that controls another, e.g., a joystick or kinematically-similar replica of the robot.
[2] Teleoperation is operation of a machine or robot at a distance.

X. Wang (Ed.): Mixed Reality and Human-Robot Interaction, ISCA 47, pp. 35–51.
springerlink.com

(a) (b)

Fig. 1. Argonne's Dual Arm Work Platform consisting of (a) remote robot and cameras, and (b) local user interface

Argonne's experience with teleoperation – and its difficulties – is not uncommon. Menchaca-Brandan and colleagues (2007) refer to similar struggles encountered during training of NASA astronauts. Teleoperated robots are now regularly used in space and underwater explorations (e.g., Menchaca-Brandan et al., 2007), military reconnaissance, and nuclear servicing (e.g., Park et al., 2002). These applications represent billions of dollars of funding and liability – for example, the liability for nuclear decommissioning work in the United States alone is around $30 billion (Park et al., 2004). Furthermore, teleoperation allows humans to reach into environments in ways that are otherwise impractical, such as in urban search and rescue (e.g., Murphy, 2004) or in minimally invasive surgery (e.g., Guthart and Salisbury, 2000).

Much of the mental burden in teleoperation, and human-robot interaction in general, is the result of misalignment in the hand-eye control of the robot. For example, to move a teleoperated robot an operator must determine the desired motion as seen in the camera views, and then impart the corresponding force or motion at the manipulandum. Thus, the operator must mentally transform (i.e., rotate, translate, scale, or deform) desired robot actions into necessary control inputs.

This chapter discusses the mental transformations found in human-robot interaction by investigating those inherent in teleoperation. It reviews relevant background, analyzes interfaces with only one camera and display, analyzes interfaces with multiple cameras and displays, and then examines the design of a low-mental-transformation interface.

The goal of this research is different from much of the literature. Rather than *adding* information into the interface to make the tasks easier (such as through augmented reality), it aims to reduce the complexity of the task by *removing* the need for mental transformations. If human-robot interfaces are made mentally efficient, then trial-and-error control may be reduced, training and task times decreased, and task performance improved.

2 Background

2.1 Components and Coordinate Frames

In general, teleoperation involves two worksites: a local worksite with the manipulandum, user interface, and human user, and a remote worksite with the robot (fixed or mobile) and several cameras (and possibly other sensors). The cameras may be on the robot or external to it, and may have tilt, pan, or zoom capabilities. The cameras' images are shown on displays in the user interface; for simplicity, assume each camera corresponds to one display.

Coordinate frames can be defined for each of the components – see Figure 2 and (DeJong et al., 2004). At the remote site there are:

- The site's world coordinates, $\boldsymbol{W}_R$
- The robot's control coordinates, $\boldsymbol{R}$
- Camera *i*'s view coordinates, $\boldsymbol{C}_i$

At the local site, there are:

- The site's global coordinates, $\boldsymbol{W}_L$
- The manipulandum's control coordinates, $\boldsymbol{M}$
- Display *i*'s view coordinates, $\boldsymbol{D}_i$
- The human operator's view coordinates, $\boldsymbol{H}$

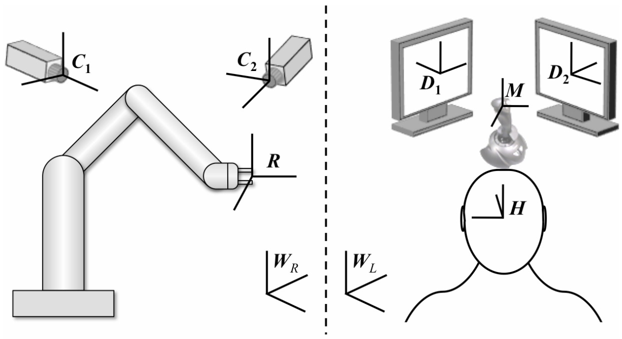


Fig. 2. Sample teleoperation components and their coordinate frames

Coordinate frames $\boldsymbol{M}$ and $\boldsymbol{R}$ represent the control of the robot – here, an input along one axis in $\boldsymbol{M}$ moves the robot along a corresponding axis in $\boldsymbol{R}$. In practice, $\boldsymbol{R}$ may be defined in a variety of ways, such as at the base of the robot or aligned with a tool, and it may be mobile or fixed. $\boldsymbol{M}$ may not correspond to the manipulandum's physical frame. In some situations, such as with joint-controlled robots, these frames are more complicated, although this analysis holds.

Each camera and display can be interpreted as a bounded plane (for the camera this is the image plane; for the display this is the screen's plane) and a perpendicular line through the center of that plane. This centerline represents the angle a camera or display is pointing. Since $\boldsymbol{C}_i$'s positive centerline points out of the camera, $\boldsymbol{D}_i$'s positive centerline points into the display. $\boldsymbol{C}_i$ may be fixed or moving if the camera has tilt/pan/zoom capabilities.

The human frame $\boldsymbol{H}$ is located at the eyes of the operator and moves with the eyes when the user looks around. One axis is where the eyes are looking; the point of focus lies on this line. The remaining two axes are in the peripheral directions.

In most teleoperation situations, the location transformations can be measured or calculated for each of the components. For example, using Special Euclidian notation (SE(3)), the transformation from remote world frame to the robot control is

$$\boldsymbol{R} = {}^{\boldsymbol{R}}_{\boldsymbol{W}_R}\mathbf{T} \cdot \boldsymbol{W}_R = \begin{bmatrix} \begin{pmatrix} & & \\ & {}^{\boldsymbol{R}}_{\boldsymbol{W}_R}\mathbf{R} & \\ & & \end{pmatrix} & \begin{pmatrix} \\ {}^{\boldsymbol{R}}_{\boldsymbol{W}_R}\mathbf{t} \\ \\ \end{pmatrix} \\ 0 \quad 0 \quad 0 & 1 \end{bmatrix} \cdot \boldsymbol{W}_R \,, \tag{1}$$

where ${}^{\boldsymbol{R}}_{\boldsymbol{W}_R}\mathbf{R}$ and ${}^{\boldsymbol{R}}_{\boldsymbol{W}_R}\mathbf{t}$ are the rotation and translation matrices from $\boldsymbol{W}_R$ to $\boldsymbol{R}$. In the same fashion, matrices ${}^{\boldsymbol{C}_i}_{\boldsymbol{W}_R}\mathbf{T}$, ${}^{\boldsymbol{D}_i}_{\boldsymbol{W}_L}\mathbf{T}$, ${}^{\boldsymbol{M}}_{\boldsymbol{W}_L}\mathbf{T}$, and ${}^{\boldsymbol{H}}_{\boldsymbol{W}_L}\mathbf{T}$ are the transformations from the corresponding world frame to camera *i*, display *i*, manipulandum, and human. For more information regarding rotational and translational matrices, see (Mason, 2001).

These location matrices can be combined into more useful component-to-component transformations, such as from manipulandum to robot,

$${}^{\boldsymbol{R}}_{\boldsymbol{M}}\mathbf{T} = {}^{\boldsymbol{R}}_{\boldsymbol{W}_R}\mathbf{T} \cdot {}^{\boldsymbol{W}_R}_{\boldsymbol{W}_L}\mathbf{T} \cdot {}^{\boldsymbol{W}_L}_{\boldsymbol{M}}\mathbf{T} \,, \tag{2}$$

or from human to display,

$${}^{\boldsymbol{D}_i}_{\boldsymbol{H}}\mathbf{T} = {}^{\boldsymbol{D}_i}_{\boldsymbol{W}_L}\mathbf{T} \cdot {}^{\boldsymbol{W}_L}_{\boldsymbol{H}}\mathbf{T} \,. \tag{3}$$

However, the camera-to-display transformations, ${}^{\boldsymbol{D}_i}_{\boldsymbol{C}_i}\mathbf{T}$, cannot be directly calculated from the camera and display location matrices, because each display image is a two-dimensional projection of the three-dimensional workspace. Assume that theses projections are linear.[3] Points in $\boldsymbol{C}_i$ are projected onto the camera's bounded image plane, and shown on $\boldsymbol{D}_i$'s bounded screen plane, often at a different scaling. This projection, shown in Figure 3 where f_i is the camera *i'*s focal length, is a scaling dependent on distance from the camera. That is, by similar triangles, a point (u_i, v_i, w_i) in the camera's frame is flattened to

$$(u_i, v_i, w_i) \rightarrow \frac{f_i}{w_i + f_i} \cdot (u_i, v_i, 0) \tag{4}$$

[3] This assumption does not hold for wide-angle cameras (Weng et al., 1992).

on the image plane. The boundedness of the image plane and display mean that only a limited range of points are mapped from one frame to the other. Therefore, the transformation from camera to display can be interpreted as a camera projection scaling, a camera image plane cropping, a display scaling (for the display's resolution, pixel size, and aspect ratio), finally followed by a display cropping to fit the screen.

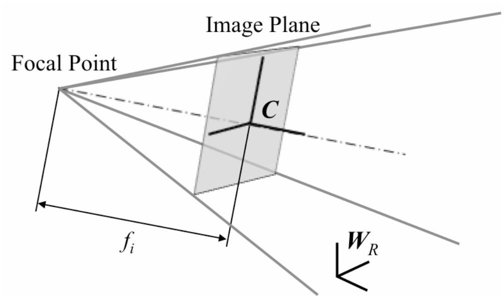


Fig. 3. Projection of points in the camera coordinate frame

2.2 The Cost of Mental Tranformations

Human-robot interfaces frequently have a misalignment between inputs at the manipulandum and corresponding robot actions, as seen in the displays. These misalignments force the user to mentally transform between control outputs and inputs.

We encounter such control misalignment quite frequently in our daily lives, although it is usually simple and easy to learn. For example, controlling a computer mouse consists of sliding a horizontal mouse to move a vertical pointer – the task is (initially) more difficult when switching to a new computer with a different mouse scaling or if the mouse frame is rotated with respect to the screen frame. Driving a car consists of rotating a vertical steering wheel to steer our horizontal path – the task becomes more difficult when driving in reverse. First-person and third-person video and computer games often have inverted pitch or yaw – in third-person games, the task is temporarily harder when the character is facing sideways or backwards in the camera view. Remote controlled cars, boats, and planes are steered in their local coordinates that may be rotated from the operator’s frame – the task is noticeably difficult when the vehicle is driving towards the operator.

While we learn simple transformations quite readily, research shows that they can be mentally taxing, especially when first learning.

In the cognitive science realm, studies show that all mental transformations can be challenging. Shepard and Metzler (1971) found that the time to mentally rotate objects is linearly related to the angle of rotation. Wexler et al. (1998) studied the relationship between mental and motor rotation. Anderson (1995) and Kosslyn (1980) performed experiments in mental image scanning showing that the time to mentally translate is related to distance. The results from these sources and others (e.g., Pylyshyn, 1979; Barfield et al., 1988) show that mental rotation is by far the most costly.

Also in the cognitive science literature, there is much discussion on mental workload[4] (Hancock and Chignell, 1988; Schvaneveldt et al., 1998; Johnson et al., 2003; DiDomenico, 2003). Most of the research has focused on reducing the mental workload of multiple sensor feedback. Since it has been shown in literature that mental transformations are costly, it follows that reducing them will decrease mental workload and improve task performance.

Furthermore, there has been research on hand-eye coordination and adaptation to errors in it. Miyake (1982) tested rotational misalignment for various arm positions and found that motions were identified in egocentric, rather than external, coordinates. Many studies have addressed visual distortions, such as nonlinear mappings by Flanagan and Rao (1995) and rotations by Krakauer et al. (2000).

Even in teleoperation literature, proper hand-eye coordination is not a new topic. For example, Sheridan (1992) discusses it in terms of a misalignment of proprioception, or "sense of self". Menchaca-Brandan et al. (2007) ran an experiment showing that users' spatial abilities correlate to task performances. A different teleoperation experiment (DeJong et al., 2004) shows that interface arrangement can affect task time and accuracy. Other experiments show that task performance depends on the definition of robot's control frame (Hiatt and Simmons, 2006; Lamb and Owen, 2005). For fixed-base robots with a single camera, control alignment is often achieved by computationally rotating the manipulandum's frame, as done by Thompson (2001), or in the well-known DaVinci surgical system (Guthart and Salisbury, 2000). Regarding multi-camera systems, Chiruvolu et al. (1992) show that hand-eye misalignments are even more critical when the operator is simultaneously using multiple video displays.

Many human-robot interfaces do not attempt to reduce mental transformations, but rather try to make them easier by adding augmented reality overlays. Cao (2000) presents the use of augmented reality navigational and orientational cues in endoscopy, to improve navigation and help surgeons "effectively perform mental rotations and mappings". Nawab et al. (2007) overlay their robot's frame onto the displays, to help the user learn the transformations. Similarly, Nielsen and his colleagues (2007) use augmented reality to fuse video, map, and robot-pose information, to help the user "mentally correlate the sets of information".

3 Single Camera/Display Interface

Given the previous terminology, this section examines the mental transformations found in a teleoperation setup with only one camera and display. In general, a mental transformation is required any time the perceived robot[5] frame is misaligned with the human operator's internal frame. There are three primary sources of

[4] Sanders and McCormick (1993) define mental workload as "a measurable quantity of the information processing demands placed on an individual by a task."

[5] The perceived robot is the robot as seen in the video displays.

misalignment: rotation between perceived robot and manipulandum (called control rotation), translation between perceived robot and manipulandum (called control translation), and rotation between human and display (called view rotation).

3.1 Control Rotation

The relationship that proves the most critical is the rotation between manipulandum and the perceived robot, called *control rotation*. Suppose the manipulandum and perceived robot are those shown in Figure 4: the perceived $\boldsymbol{R}$ is rotated with respect to $\boldsymbol{M}$. To move the robot in one direction, the operator may need to push or move the manipulandum in an entirely different direction. This mental rotation can be confusing, especially if it is about an axis that is neither vertical nor horizontal, or if it is a large rotation. To have no control rotation, the transformation from manipulandum to robot must be identity, $\boldsymbol{I}$:[6]

$$ {}^{R}_{M}\mathbf{R} = {}^{R}_{C}\mathbf{R} \cdot {}^{C}_{D}\mathbf{R} \cdot {}^{D}_{M}\mathbf{R} = \boldsymbol{I} \ . \tag{5} $$

Recall that the transformation from camera to display is a scaling and cropping (i.e., no rotation), so ${}^{C}_{D}\mathbf{R} = \boldsymbol{I}$. Therefore, to eliminate control rotation, the rotation from robot to camera must be the same as that from manipulandum to display:

$$ {}^{C}_{R}\mathbf{R} = {}^{D}_{M}\mathbf{R} \cdot \tag{6} $$

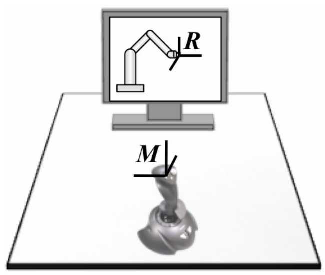


Fig. 4. Single camera and display interface with control rotation

Often, a camera used in teleoperation can pan or tilt, meaning that its orientation matrix, ${}^{C}_{R}\mathbf{R}$, changes. To avoid control rotation when panning or tilting the camera, the robot control frame, the manipulandum control frame, or the display itself must be rotated in unison. Rotating the display affects the human/display relationship, as shown later, and has limited range before the image is physically unviewable. Rotating the robot frame may destroy an intuitive control relationship with regards to the robot's kinematics. For example, if the robot frame was chosen

[6] The subscript *i* has been dropped because there is only one camera and display.

such that one of the axes is aligned with an arm of the robot, the rotated frame may alter this relationship.

Therefore, computationally or physically rotating the manipulandum control frame is usually the best choice, although it still has dangers. If the user is holding the manipulandum while panning or tilting the camera, then rotating the manipulandum's frame can cause erroneous input. Computational rotation is simple to implement and requires no additional hardware, but it may destroy kinematic mapping if the manipulandum is kinematically similar to the robot. On the other hand, physical rotation requires additional hardware, but it offers visual and haptic feedback of the rotation.

3.2 Control Translation

Even if the control is rotationally aligned, it may still require mental translations, called *control translation*. Suppose the manipulandum and the perceived robot are those in Figure 5, where the frames are oriented the same, but the manipulandum is translated to the side. The user needs to mentally translate control inputs from the perceived robot to the manipulandum. For example, if the desired motion of the robot is up-right-back, then the required input is on a parallel line but translated to the manipulandum. To eliminate control translation, the translation matrices must be identical:

$$ {}^{C}_{R}\mathbf{t} = {}^{D}_{M}\mathbf{t} \,. \tag{7}$$

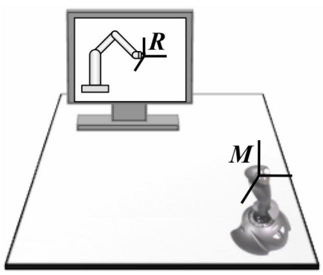


Fig. 5. Single camera and display interface with control translation

As shown in the literature, mental translation is nontrivial, but not as burdensome as mental rotation. In fact, almost all human-robot interfaces involve some control translation, such as along the depth axis of the display (i.e., the perceived robot in a physical display cannot be in the exact same location as a physical manipulandum). For that matter, depth in the display is hard to quantify, because the image is a projection onto the image plane. The user may perceive the robot as near (or far) if it appears larger (or smaller) than expected.

Panning or tilting the camera affects control translation as it did control rotation. Panning and tilting move the camera about a center of rotation that is usually not the center of the image plane, so the image plane is rotated and translated with respect to world coordinates.[7] Thus, the perceived robot translates and rotates in the video image.

In addition, camera zoom affects the control relationship, although it is less significant. Zooming a camera (i.e., changing its focal length) changes the scaling in its projection: points in space move radially to and from the image's center. That is, the perceived robot changes size and location in the display's image. Therefore, camera zoom (ignoring any cropping effects) is similar to a three-dimensional translation of the robot, camera, display, or manipulandum.

3.3 View Rotation

The third transformation of interest is the view transformation from human user to display, ${}^{D}_{H}\mathbf{T}$. The human coordinate frame is defined as always pointing in the direction the user is looking, such as at the display. This definition restricts human motions to three dimensions: longitudinal and lateral rotation about the display (as on a sphere), and translation towards and away from the display (radially).

Translation of the human frame towards and away from the display (radially), without changing the control relationship, does not require any additional mental transformations. Consideration should be given, of course, to the viewable distance of the display and its resolution. This also means that the size of the display does not affect mental transformations, as long as the display is not so large that the user cannot view all of it without turning his head.

On the other hand, rotation of the human frame about the view (longitudinally and laterally) is important. This rotation is called *view rotation*. Suppose the human and display are situated as in Figure 6, with the human frame rotated from the display's centerline. Since the image on the display is two dimensional, the user sees the same information regardless of angle. The angled image is not of the perceived robot from a new angle as it would be if the user moved about the real robot. Because of the user's mental model of the monitor and the depth into the image, the image is interpreted (correctly) as a rotated two-dimensional plane. To move the robot, the user now must determine the desired robot action in this rotated plane, requiring mental rotation of the image[8] so they are perpendicular. That is, the mental model of the planar image is mentally rotated to determine control inputs. To eliminate view rotation, the human's view axis must be perpendicular to the displays plane, *at the area of interest*. This constraint is clearly impractical because it means that as the user changes his area of interest, he must translate. Thus, a more reasonable constraint is to place the human on the display's centerline to reduced view rotation. This constraint can be represented mathematically by the conditions

---

[7] If the center of rotation is the center of the image plane, then the translation component is zero.

[8] Or the user, though this appears to be harder (Wraga et al., 1999).

$$ {}^{D}_{H}\mathbf{R} = I \text{ and } {}^{D}_{H}\mathbf{t} = \begin{pmatrix} 0 \\ 0 \\ l_0 \end{pmatrix}, \tag{8} $$

where l_0 is the radial distance from human to display.

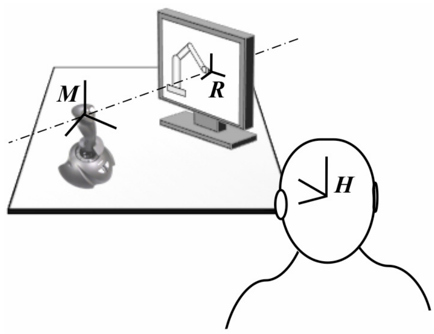


Fig. 6. Single camera and display interface with view rotation

For view rotation, using stereovision is more harmful than using monoscopic vision because of “induced stereo movement” (Tyler, 1974; Arthur; 1993). If the display is stereoscopic, the image appears three-dimensional when the user is looking perpendicular at the screen. Imagine that the image on the screen is a hand pointing at the user. When the user is rotated to the side, the two stereoscopic images are the same, so the hand still appears to be pointing at the user rather than in the original direction. With teleoperation, this means that the perceived robot has rotated with respect to the manipulandum. Thus, with stereovision, it is necessary that the human remains on the display's centerline. Systems like the DaVinci telesurgical robot that rely heavily on stereovision restrict the human’s head to a fixed location and orientation (Guthart and Salisbury, 2000).

3.4 How to Design a Single Camera/Display Interface

Therefore, there are three relationships of interest, in order of importance: control rotation, view rotation, and control translation. The cognitive science literature shows that these transformations are mentally taxing, and the human-robot literature shows that reducing them can increase task performance (DeJong, 2003).

In designing an interface with only one camera and one display, the designer may wish to position the camera, manipulandum, and video display properly so as to eliminate or minimize the control transformations. Given the locations of any two of these three components, the designer can calculate the location for the third component that follows Eqs. 6-7. For example, given the camera and display locations, there is a specific position and orientation for the manipulandum that minimizes the mental transformations.

Once the control is aligned, the user should be situated on the display's centerline, to follow Eq. 8.

4 Multiple Camera/Display Interfaces

Often, human-robot interfaces incorporate more than one camera and display to give the user additional visual feedback of the remote worksite. The previous single camera/display methods are easily extended to interfaces with multiple cameras and displays.

4.1 Multiple Control and View Relationships

When an interface involves more than one camera-display pair, the control and view alignments are especially important. If the user is using one of the displays and it involves mental transformations, he may eventually learn the mapping such that control is relatively easy. However, once the user switches attention to a different display, the mapping has changed, and he must learn a new mapping. Even when switching back to the original display, the user must relearn the relationships. This switching and relearning has been shown to be mentally taxing (Chiruvolu et al., 1992).

For example, suppose the interface consists of two displays, as shown in Figure 7. This interface has properly aligned manipulandum and perceived robot frames for the camera-display on the left. When the user is using the left display, control is straightforward: an input on the manipulandum moves the perceived robot in the same way. However, when the user attempts to use the display on the right, the control is misaligned: an input moves the perceived robot in an unintuitive way.

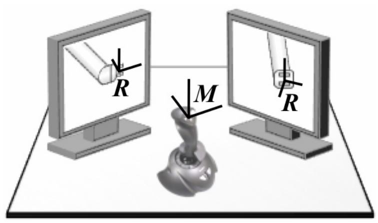


Fig. 7. An example of a mentally inefficient interface with two cameras and displays

4.2 How to Design a Multiple Camera/Display Interface

Designing a mentally efficient interface is more complicated when it involves multiple control and view mappings. Each camera/display should independently satisfy the control and view constraints mentioned previously.

For simplicity, assume that the cameras and robot frames are fixed and known, but the displays, manipulandum, and human frames need to be arranged. This is often the case. The robot frame is often intelligently chosen for the given application and robot kinematics. Similarly, the cameras are often chosen based on where they physically can be placed and where they are most beneficial for the given tasks.

The best solution is to align the manipulandum with every perceived robot frame simultaneously, by placing the manipulandum and then arranging the displays around it such that Eq. 6 is satisfied for each display. If this can be done, then the control is aligned for whichever display the user is currently using – more importantly, there is no need to learn a new mapping when switching displays. Figure 8 shows a two-display teleoperation setup that achieves this. This method arranges the displays on a sphere around the user and manipulandum, such as that in Figure 9.

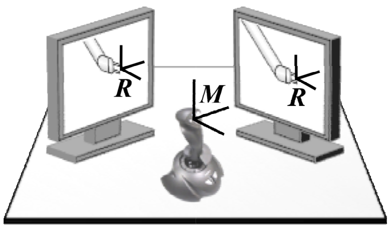


Fig. 8. An example of an aligned interface with two cameras and displays

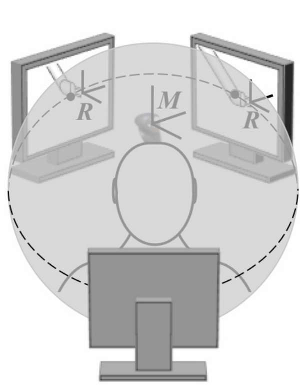


Fig. 9. Method of placing displays on sphere around the user

A constraint that follows from simultaneous alignment is that the relative orientation of the displays must be the same as the relative orientation of the cameras. For simultaneous rotational alignment of any two displays with the same robot and manipulandum (recall Eq. 6),

$$ {}^{C_1}_{R}\mathbf{R} = {}^{D_1}_{M}\mathbf{R} \text{ and } {}^{C_2}_{R}\mathbf{R} = {}^{D_2}_{M}\mathbf{R} \,, \tag{9} $$

which leads to:

$$ \begin{gathered} {}^{C_1}_{R}\mathbf{R}\cdot{}^{R}_{C_2}\mathbf{R} = {}^{D_1}_{M}\mathbf{R}\cdot{}^{M}_{D_2}\mathbf{R} \\ {}^{C_1}_{C_2}\mathbf{R} = {}^{D_1}_{D_2}\mathbf{R} \end{gathered} \tag{10} $$

This makes sense: if the two cameras are perpendicular to each other, the displays must be perpendicular as well. The camera on the left as viewed from the robot (facing the front of the camera) corresponds to the display on the left as viewed by the user (facing the front of the display).

This ideal solution is not always feasible, such as when the aligned location for a display is not physically achievable due to space conflicts. When the displays cannot be aligned simultaneously, the issue becomes how to enforce alignment for the display the operator is currently using.

One way to accomplish alignment is to perform a transition of the manipulandum frame when the user is switching views. In Figure 7, this means that when the user switches attention to the display on the right, the manipulandum's frame must be rotated (computationally or physically) to align it with the new perceived robot frame. The transition can be initiated from various sources, such as eye tracking or the user pressing a key corresponding to the new display.

Unfortunately, transitioning the manipulandum frame has drawbacks. First, rotating it dynamically can cause erroneous motion. Conversely, forcing the user to stop inputs during transition increases task time. Second, passive measures of attention switching, such as eye tracking, may initiate transition when the user is merely glancing between displays. On the other hand, requiring the user to press a button when switching adds time and may increase overall mental workload for an interface. Third, if the task requires close attention to both views simultaneously (e.g., when trying not to collide with multiple objects seen in separate views), transition becomes completely infeasible.

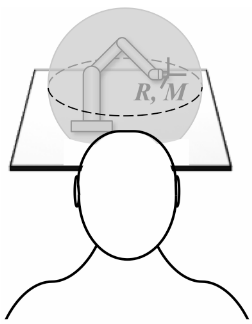


Fig. 10. Method of placing displays facing out from a common point – the ideal of which is a hologram

An alternative method to placing the displays around the user is to place the user and manipulandum around the displays. If the displays can be placed facing out from a central point, then the user and manipulandum can be moved around the displays while keeping control and view alignment. This method applies if the sphere of displays is replaced with a hologram of the robot, in which case a kinematically similar manipulandum can be placed physically coincident with the perceived robot (see Figure 10).

4.3 Ergonomics and Camera Motion

One disadvantage to the arranging the displays as presented in the previous section is that the location and orientation for a display that minimizes mental transformations may be in an ergonomically poor location. It is possible that a display should be placed above or even behind the operator to minimize transformations. For example, if two cameras face in opposite directions (e.g., front and back or left and right of the robot), then the displays should also do so (as in Figure 9).

5 Argonne Redesigned

Motivated by Argonne's past experience with teleoperation, we (the authors and our colleagues, including Argonne researchers) established a collaborative testbed for improving teleoperation. The remote worksite was located at Argonne and included a six-degree-of-freedom robotic arm, two video cameras, and a structured light system (Park et al., 2004). The local worksite was located thirty miles away at Northwestern University and consisted of a cobotic manipulandum (Faulring et al., 2006) and two displays with augmented reality. The manipulandum, robot, and user interface communicated via Ethernet. The testbed successfully implemented reduced mental transformations, structured light sensing, virtual surfaces, cobotic technology, and augmented reality aids (DeJong et al., 2006).

In designing the interface, we carefully arranged the components to minimize mental transformations. We positioned the displays around the user with simultaneous control alignment using the following steps (see Figure 11).

First, we placed the cameras at the remote site to provide useful views of the robot's workspace.

Second, we placed the two displays in front of the user such that the user was at the intersection of their centerlines. Doing so minimized view rotation without transition needed when shifting from one view to the other. Furthermore, we angled the displays to match the angle between the cameras (following Eq. 10).

Third, we needed to properly position the manipulandum to minimize control rotation and translation. Our manipulandum was not kinematically similar to the robot, so we had freedom in the computational definition of the manipulandum frame. Because of the size and shape of our cobotic manipulandum, we placed it on a table between our two displays. Doing so also minimized translational misalignment between manipulandum and perceived robot, i.e., it minimized control translation.

Finally, we needed to orientate the manipulandum frame to eliminate rotational misalignment between manipulandum and perceived robot, i.e., eliminate control rotation. The manipulandum had an internal reference frame for recording inputs, $\boldsymbol{Int}$, and a computational definition of the manipulandum control frame, ${}^{\boldsymbol{M}}_{\boldsymbol{Int}}\mathbf{R}$. Using one camera and display, we calculated this definition from Eq. 6, knowing ${}^{\boldsymbol{R}}_{\boldsymbol{C}}\mathbf{R}$ and ${}^{\boldsymbol{D}}_{\boldsymbol{Int}}\mathbf{R}$:

$$ {}^{\boldsymbol{M}}_{\boldsymbol{Int}}\mathbf{R} = {}^{\boldsymbol{M}}_{\boldsymbol{D}}\mathbf{R}\cdot{}^{\boldsymbol{D}}_{\boldsymbol{Int}}\mathbf{R} = {}^{\boldsymbol{R}}_{\boldsymbol{C}}\mathbf{R}\cdot{}^{\boldsymbol{D}}_{\boldsymbol{Int}}\mathbf{R} \, . \tag{11} $$

The manipulandum recorded inputs, multiplied them by this matrix, and then commanded them to the robot.

The teleoperation testbed showed significant improvement in task performance (DeJong et al., 2006). When arbitrary and minimized-transformation interfaces were both used by Argonne personnel, they were surprised by how much the reduced transformations made human-robot interaction easier. Figure 11 shows the mentally efficient interface.

Fig. 11. Mentally efficient interface for Northwestern-Argonne teleoperation

6 Conclusion

Clearly, mental transformations in human-robot interfaces degrade task performance, increase the skill required by users, and are mentally tiresome for the user. These transformations come from misalignment between manipulandum and perceived robot (control rotation and translation), and user and display (view rotation).

Ideally, human-robot interfaces should be carefully designed to minimize the mental transformations. This can be accomplished through intelligent arrangement of the interface's components, as shown in the sample implementation. Doing so makes tasks inherently easier, and can increase task performance.

Acknowledgements

The authors would like to thank their Argonne colleagues Young Park, Hyosig Kang, and Thomas Ewing, as well as Eric Faulring for use of the cobotic manipulandum. The original research was funded by the Department of Energy.

References

Anderson, J.R.: Review: perception-based knowledge representations. In: Cognitive Psychology and Its Implications, 4th edn. Freeman, New York (1995)

Arthur, K.W.: 3D task performance using head-coupled stereo displays. M.S. thesis, University of British Columbia (1993)

Barfield, W., Sandford, J., Foley, J.: The mental rotation and perceived realism of computer-generated three-dimensional images. International Journal of Man-Machine Studies 29(6), 669–684 (1988)

Cao, C.G.L.: Augmented reality displays for endoscopic orientation and navigation. In: CHI 2000 extended abstracts on Human Factors in Computing Systems (2000)

Chiruvolu, R., Hwang, V.S., Sheridan, T.B.: Virtual display aids for teleoperation. In: SPIE – The International Society for Optical Engineering (1992)

DeJong, B.P.: Improving Teleoperation Interfaces: Creation of a Testbed and Thinking Inside the Box. M.S. thesis, Northwestern University (2003)

DeJong, B.P., Colgate, J.E., Peshkin, M.A.: Improving teleoperation: reducing mental rotations and translations. In: IEEE International Conference on Robotics and Automation, New Orleans (2004)

DeJong, B.P., Colgate, J.E., Peshkin, M.A.: Improving teleoperation: reducing mental rotations and translations. American Nuclear Society, Gainesville, FL (2004)

DeJong, B.P., Faulring, E.L., Colgate, J.E., Peshkin, M.A.: Lessons learned from a novel teleoperation testbed. Industrial Robot 33(3), 187–193 (2006)

Department of Energy: Dual Arm Work Platform Teleoperated Robotics System, DOE/EM-0389 (1998)

DiDomenico, A.T.: An investigation on subjective assessments of workload and postural stability under conditions of joint mental and physical demands. Ph.D. thesis, Virginia Polytechnic Institute and State University (2003)

Faulring, E.L., Colgate, J.E., Peshkin, M.A.: The cobotic hand controller: design, control and performance of a novel haptic display. The International Journal of Robotics Research 25, 1099–1119 (2006)

Flanagan, J.R., Rao, A.K.: Trajectory adaption to a nonlinear visuomotor transformation: evidence of motion planning in visually perceived space. Journal of Neurophysiology 74(5), 2174–2178 (1995)

Guthart, G.S., Salisbury, J.K.: The intuitive telesurgery system: overview and application. In: IEEE International Conference on Robotics and Automation (2000)

Hancock, P.A., Chignell, M.H.: Mental workload dynamics in adaptive interface design. IEEE Transactions on Systems, Man, and Cybernetics 18(4), 647–658 (1988)

Hiatt, L.M., Simmons, R.: Coordinate frames in robotic teleoperation. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (2006)

Johnson, C.A., Adams, J.A., Kawamura, K.: Evaluation of an enhanced human-robot interface. In: IEEE International Conference on Systems, Man, and Cybernetics (2003)

Kosslyn, S.M.: Image and Mind. Harvard University Press, Cambridge (1980)

Krakauer, J.W., Pine, Z.M., Ghilardi, M.F., Ghez, C.: Learning of visuomotor transformations for vectorial planning of reaching trajectories. Journal of Neuroscience 20(23), 8916–8924 (2000)

Lamb, P., Owen, D.: Human performance in space telerobotic manipulation. In: ACM Symposium on Virtual Reality Software and Technology (2005)

Mason, M.T.: Mechanics of Robotic Manipulation. MIT Press, Cambridge (2001)

Menchaca-Brandan, M.A., Liu, A.M., Oman, C.M., Natapoff, A.: Influence of perspective-taking and mental rotation abilities in space teleoperation. In: ACM/IEEE International Conference on Human-Robot Interaction (2007)

Miyake, Y.: Coordinate systems for hand control. Ph.D. thesis, University of California (1982)

Murphy, R.R.: Human-robot interaction in rescue robotics. IEEE Transactions on Systems, Man, and Cybernetics 34(2) (2004)

Nawab, A., Chintamani, K., Ellis, D., Auner, G., Pandya, A.: Joystick mapped augmented reality cues for end-effector controlled tele-operated robots. In: Virtual Reality Conference (2007)

Nielsen, C.W., Goodrich, M.A., Ricks, R.W.: Ecological interfaces for improving mobile robot teleoperation. IEEE Transactions of Robotics and Automation 23(5) (2007)

Noakes, M.W., Love, L.J., Lloyd, P.D.: Telerobotic planning and control for DOE D and D operations. In: IEEE International Conference on Robotics and Automation (2002)

Park, Y.S., Ewing, T.F., Yule, T.J., Colgate, E., Peshkin, M.: Enhanced teleoperation exhibiting tele-autonomy and tele-collaboration. Department of Energy, ANL/TD/CP-107730 (2002)

Park, Y.S., Kang, H., Ewing, T.F., Faulring, E., Colgate, J.E., Peshkin, M.: Enhanced teleoperation for D&D. In: IEEE International Conference on Robotics and Automation, New Orleans (2004)

Pylyshyn, Z.W.: The rate of 'mental rotation' of images: a test of a holistic analogue hypothesis. Memory and Cognition 7, 19–28 (1979)

Sanders, M.S., McCormick, E.J.: Human Factors in Engineering and Design, 7th edn. McGraw-Hill, New York (1993)

Schvaneveldt, R.W., Gomez, R.L., Reid, G.B.: Modeling mental workload. Cognitive Technology 3(1) (1998)

Shepard, R.N., Metzler, J.: Mental rotation of three-dimensional objects. Science 171, 701–703 (1971)

Sheridan, T.B.: Telerobotics, Automation, and Human Supervisory Control. MIT Press, Cambridge (1992)

Thompson, R.L.: Integration of visual and haptic feedback for teleoperation. Ph.D. thesis, University of Oxford (2001)

Tyler, W.: Induced stereo movement. Vision Research 14, 609–613 (1974)

Weng, J., Cohen, P., Herniou, M.: Camera calibration with distortion models and accuracy evaluation. IEEE Transactions on Pattern Analysis and Machine Intelligence 14(10) (1992)

Wexler, M., Kosslyn, S.M., Berthoz, A.: Motor processes in mental rotation. Cognition 68, 77–94 (1998)

Wraga, M., Creem, S.H., Proffitt, D.R.: The influence of spatial reference frames on imagined object and viewer rotations. Acta Psychologica 102, 247–264 (1999)

User-Centered HRI: HRI Research Methodology for Designers

M. Kim, K. Oh, J. Choi, J. Jung, and Y. Kim

Korea Advanced Institute of Science and Technology (KAIST), Republic of Korea

Abstract. This chapter introduces the field of user-centered HRI, which differs from the existing technology-driven approach adopted by HRI researchers in emphasizing the technological improvement of robots. It proposes a basic framework for user-centered HRI research, which comprises the three elements of "aesthetic", "operational", and "social" contextuability. This framework is made for robot product designers seeking to incorporate user perspectives and needs; it is intended to allow easy identification and efficient study of issues in user-centered HRI design. The case studies introduced in this chapter, all of which are based on the aforementioned framework, will facilitate understanding of user-centered HRI, and create new research opportunities for designers, non-experts, and robot engineers.

Keywords: Human-Robot Interaction, Robot Design, User-Centered HRI.

1 Introduction

Thanks to advancements in robotics technology, the groundbreaking results of academic research, and the growth of the robotics market brought on by rising demand, robotics is moving beyond the realm of engineering and becoming an active field of research in the humanities and social sciences. Robot design, in particular, is gaining increasing importance as a means of presenting robots, commonly regarded as technocentric hardware, to new users as software products that combine communication media with services suited for everyday life.

However, the most urgent cause for vitalizing research into robot design lies in the dominant approach taken by existing researchers: because previous work on robotics aimed primarily at technological achievement, the focus tended to be almost exclusively on enhancing the skills and capabilities of robots from a developer's point of view, rather than on improving user-friendliness. For this reason, there are few products that have succeeded in winning market popularity, despite that the cognitive and decision-making capabilities of robots have reached an astonishing level. To redress this problem, active effort has begun to be expended on the study of the social attributes of robots (e.g. Fong et al., 2003; Breazeal, 2004; Dautenhan, 1999). Therefore, this chapter focuses on aspects that were lacking in the approaches taken by previous technocentric HRI studies: user-centered HRI is defined to examine the requirements of robot users as well as changes in user behavior caused by robots, and a framework for HRI research

X. Wang (Ed.): Mixed Reality and Human-Robot Interaction, ISCA 47, pp. 13–33.

based on the findings of this examination is proposed. In addition, case studies are employed to facilitate an understanding of the framework, and future possibilities for research on user-centered HRI are explored.

2 User-Centered HRI

Even when two robots are based on the same technology, they can be experienced in completely different ways by users depending on the type of interaction system they employ. For this reason, interaction design—which, together with character design and appearance design, is one of the three elements of robot design (Oh et al., 2005)—tends to be weighted above the other two elements, and interaction-related factors such as modality, autonomy, and interactivity figure prominently among the five properties of social robots mentioned by Bartneck and Forlizzi (2004). Therefore, robot designers must not only be proficient in their traditional task of appearance design, but also develop a solid grasp of the more fundamental issue of HRI.

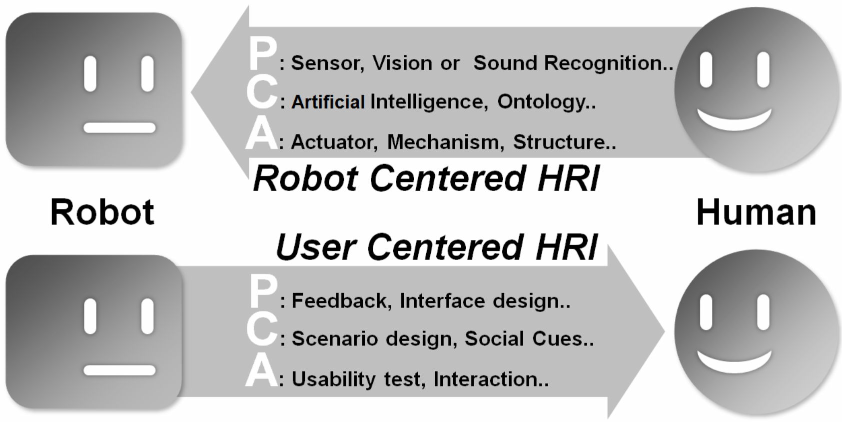


Fig. 1. Dual approaches to HRI research: robot-centered HRI and user-centered HRI

Research on human-robot interaction, or HRI, can be divided largely into two approaches: *'robot-centered HRI'* and *'user-centered HRI'* (Oh and Kim, 2007) (Figure 1). The majority of existing studies on HRI concentrate on how robots could be made to effectively perceive, cognize, and act in their given environment in order to interact well with humans. This falls under the former category of robot-centered interaction research, a perspective typically adopted by robot engineers. R. A. Brooks, C. Breazeal, and others at MIT are some of the most representative proponents of robot-centered HRI research, which focuses on robot control and implementation technologies. Unlike ICRA (International Conference on Robotics and Automation) and IROS (International Conference on Intelligent Robots and Systems), both established traditional academic conferences on robotics, RO-MAN deals only with specialized research topics relating to HRI. However, even the topics presented at this workshop, aside from rare exceptions such as "philosophical issues of robot/human coexistence," "cognitive psychology in robot/human interactions," and "intuitive instruction of robot assistants," focus primarily on how to control robots in order to enhance their recognition and perception of humans (Kim and Kim, 2005).

The second perspective, as mentioned above, is user-centered interaction research. This approach prioritizes how robot users perceive and respond to robots, and how they adjust their behavior as a result of such response, over how robots cognize and accommodate humans. In other words, user-centered HRI research can be defined as the study of interaction aimed at enabling robots to better assist human users in accomplishing their desired objectives and to provide a higher quality of user experience. Today, user-centered HRI research is becoming the province of robot designers on the grounds that designers acquire and develop specialized knowledge about how users perceive and regard products from an overall point of view (Lobach, 1997), and that they prioritize the needs and experiences of the user over those of the robot. It is also an area of research that robot engineers should explore more actively. Examples of user-centered HRI research include the study of emotional bonds or feelings that users develop in the course of interacting regularly or closely with robots, and of changes in the value of life or the ecology of other products (Forlizzi, 2007a) brought on by the infiltration of robots into everyday existence.

3 HRI Research Framework for Designers

Because the objective of designers researching user-centered HRI diverges from that of robot engineers, the very definition of the problem needs to be differentiated. This means that a differentiated outline and framework for research tailored to designers must first be developed. Thus, this chapter examines the respective research elements covered by robot engineering and robot design, and integrates the two to propose a set of research elements distinctive to HRI design, i.e. the research field of user-centered HRI.

3.1 Research Elements in Robotics Engineering: P, C, A

Research elements in robotics engineering can be divided largely into three categories based on the graduate curriculum of the Robotics Institute at Carnegie Mellon University: 1) sensor technology, which allows a robot to perceive its external environment; 2) AI technology, which enables a robot to cerebrate using the wide-ranging information and knowledge databases thus perceived, and to autonomously control its various components; and 3) actuator technology, which endows a robot with mobility and the capacity for action. Also conducted is comprehensive research on the mathematical correlation of these three elements.

When these research elements are mapped onto the cognitive process humans carry out in response to their environment (Norman, 1988; Card et al., 1983), sensor technology corresponds to ***'perception'***, AI technology to ***'cognition'***, and actuator technology to ***'action.'*** When an intelligent robot equipped with such a system communicates with human beings, the model for this communication can be schematized as shown in Figure 2 (Kim and Kim, 2005).

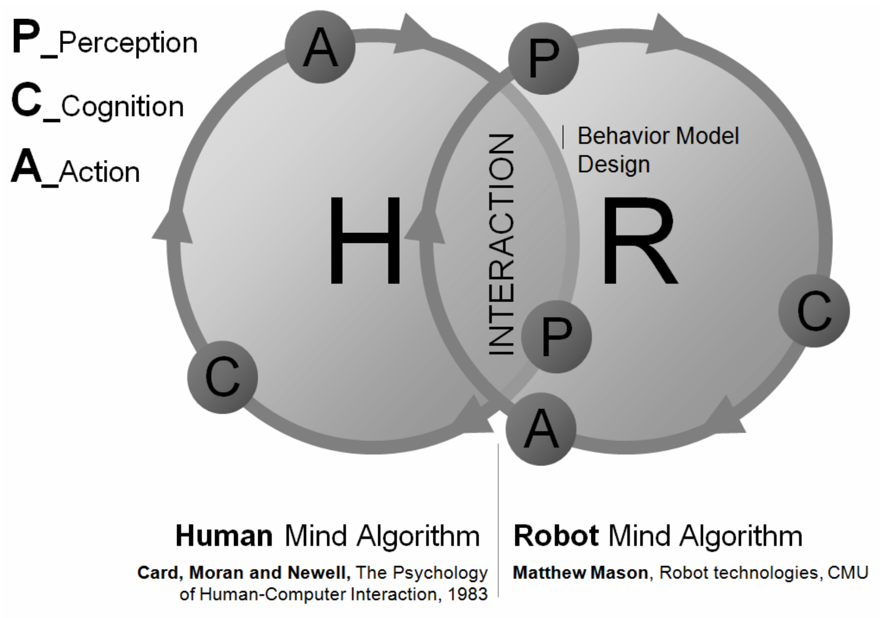


Fig. 2. P.C.A model for HRI

In the above model, the human user and the robot, each possessing the ability for P (perception), C (cognition), and A (action), follow the PCA cycle as they alternate between independent and interdependent action. Human-robot interaction thus occurs when the two PCA cycles overlap, resulting in communication. The zone in which this interaction event occurs can be termed the ***'human-robot interface.'***

3.2 Research Elements in Design

In order to link the research elements of robotics engineering—i.e. P, C, and A—with the research elements of product design, it is necessary to examine the background characteristics distinctive to the academic discipline of product design. According to studies by the pioneers of design, the objects of research can be divided into "the determinate" and "the indeterminate"; whereas the study of science falls under the former heading, design can be defined as the study of the indeterminate through invention and creation (Buchanan, 1995; Swann, 2002; Cross, 2001; Simon, 1969).

Rather than concentrating on the changing objects or media of design, design researchers must focus on the mental process, or flow of thought, involved in how we change or create objects and media. Richard Buchanan (1995) saw this mental process as a rhetorical one, and proposed the sequence known as the "4 orders of design," consisting of signs and images, physical objects, actions and services, and ideas and systems (Figure 3). Buchanan's attempt reflects the understanding that design research must regard its objects as more than mere objects, by contemplating their fundamental nature and characteristics.

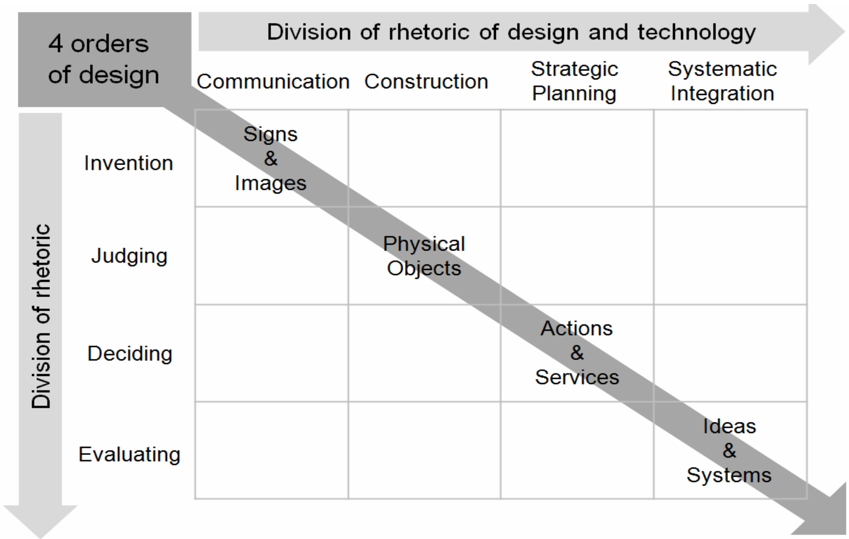


Fig. 3. Matrix of abilities and disciplines in design (Modified from Buchanan, 2001)

3.3 Research Elements in HRI Design

Each of Buchanan's four orders constitutes a specialized area of design (Figure 3). When restricted to the research elements of HRI design, "signs and images" and "physical objects" can together be mapped onto "aesthetic contextuability," while "actions and services" and "ideas and systems" can be mapped onto a robot's "operational contextuability" and "social contextuability," respectively. In short, when HRI is recast in light of the research elements of design, perception can be linked to "aesthetic contextuability," action to "operational contextuability," and cognition to "social contextuability."

First, "aesthetic contextuability" is intimately related to the information that is exchanged in the very first moments of interaction between human and robot. In terms of a user-centered approach, interaction research centering on aesthetic contextuability can be defined as the study of how users experience and respond to the external aspects of a robot's makeup, such as shape, color, material, sounds and vibrations, size and proportions, etc. which are first seen, heard, and felt through the user's five senses. In the robot design process, aesthetic contextuability is dealt with partially during the appearance design stage, but it is not restricted to a single element of robot design. Form, which constitutes the core element of aesthetic contextuability, indirectly conveys impression and character at the initiation of conversation between human and robot, and affects the determination of their method of communication. Therefore, HRI designers must not only understand form from the perspective of design, but also from that of cognitive psychology.

Some examples of such research include studies on how robots may become “human-like” as a result of certain external factors and on the “uncanny valley” hypothesis, as well as studies on formal elements suited to a robot’s specific character and role, or on what materials might best befit a robot’s given use.

Second, “operational contextuability” involves aspects of handling and manipulation, patterns of movement, and usability during actual interaction between user and robot. Such questions as how the functions served by the robot are understood and regarded by the user, and how user-robot interaction fulfills the needs of the user both fall under the heading of operational contextuability research. Depending on what function a product employs to assist the user’s task, and on what method is used to implement a given function, the user can go beyond merely having his or her expectations satisfied to experiencing pleasure (Hassenzahl, 2002).

A user’s needs are not fixed; they change according to time, place, and situation. If one can ascertain what kinds of desires can be inspired and fulfilled by using and interacting with a robot, and apply this knowledge to robot design, it will be possible to transform robots into products that can broadly impact the public’s everyday attitudes and activities, just as past innovations such as automobiles and mobile phones were able to do.

HRI research on operational contextuability starts with observing and understanding how people use robots. Although robots are operated by a user who can manipulate its controls, they are distinct from other existing products in that they perform their functions autonomously on the basis of user-robot communication or identification of environmental factors. For this reason, close examination of the perception and attitudes of users regarding the implementation of a robot’s functions is imperative for the development of superior robots.

Third and last, “social contextuability” deals with such social and metaphysical values as intimacy, sense of connectedness, significance, worth, trust, and feeling, which a user attains as a result of interacting with a robot. Human beings do much more than simply perceive, feel, and use products—particularly when it comes to products like robots, which are made to execute human functions, have conversation skills, and express emotion. Research on social contextuability thus covers modifications in role assignment caused by the placement of a robot in a human user’s everyday environment, changes in living patterns as examined in Forlizzi’s study (2007b), changes to product value including product attachment (Mugge et al., 2008), and so on.

Characteristics such as those mentioned above are not so much inherent in robots themselves, as produced through the intermingling of past knowledge, experience, and cultural habits during a user’s interaction with a robot. Therefore, studying these social and metaphysical characteristics requires an understanding of the sociocultural context surrounding users and robots, as well as an anthropological and qualitative approach to research.

The three research elements of HRI design examined thus far are aligned with the three levels of design (visceral, behavioral, and reflective) mentioned by Norman (2004), and with the three ways (aesthetically, functionally, symbolically) in which people collect and use products, as mentioned by Forlizzi (2007a). This classification, however, remains basic at best; it would be impossible to divide all research on HRI design accurately into these three categories. Still, it is possible to position studies covering aspects of these research elements in varying proportions upon a continuous triangular plane whose three axes are constituted by the three elements (Figure 4).

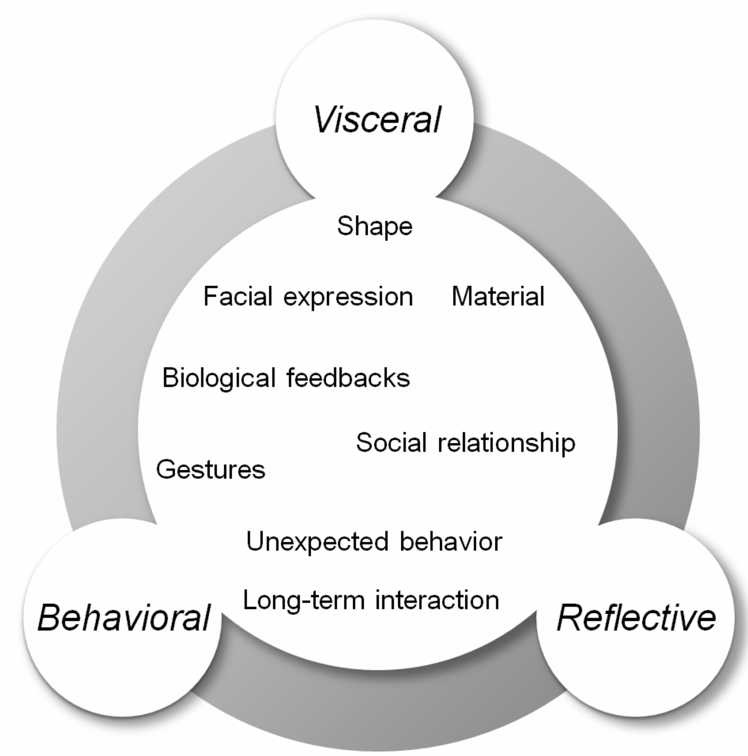


Fig. 4. The research elements of HRI design: Keywords in the inner circle imply main issues of following case studies (Chapter 4)

3.4 Research Framework for HRI Design

Because creating a robot requires enormous investments of time and money, prototyping methods have been proposed as a way to make robot research easier and more efficient for general researchers including designers, as opposed to robot engineers (Bartneck, 2004).

User-centered HRI research, unlike robot engineering, does not seek to discover specific functions explaining how HRI affects the perception and behavior of robots and users. Rather, it aims at analyzing user perception and its resultant responses, as well as the user requirements that ensue. This means that by clearly specifying the research objectives and effectively setting the independent variables for research, HRI studies can be carried out successfully even with prototypes of relatively low fidelity.

One way to study the three research elements covered in Section 3.3 is to employ the three kinds of prototypes mentioned by Yang (2004). First, to study aesthetic contextuability, it is necessary to manufacture a “look and feel” prototype. This prototype needs to recreate sensory stimuli closely approximating those provided by the actual robot, including visuals and sound.

Second, to study operational contextuability, a “function prototype” needs to be created. A function prototype is a prototype made to perform functions similar to those of the actual robot; it thus needs to be able to execute the functions intended by the designer and identify relevant user behavior. Type-To-Speech (TTS) and Wizard of Oz (WOz) are two techniques that can be used for this purpose (Gould et al., 1983). This type of prototype allows for research on task execution, such as human-robot collaboration and puzzle solving.

To study the third element of social contextuability, “role prototypes” are needed. This type of prototype is used to acquire an understanding of usage context, through the utilization of storyboarding or scenario techniques. Social contextuability is a relatively more difficult field to research than the above two, and can only be studied through insightful analysis based on long-time observation of human-robot interaction. However, when the goal and scope of research is restricted to a clear-cut, peripheral problem, using a role prototype to have a user virtually experience and imagine usage situations can allow a researcher to identify user responses or requirements in terms of social contextuability.

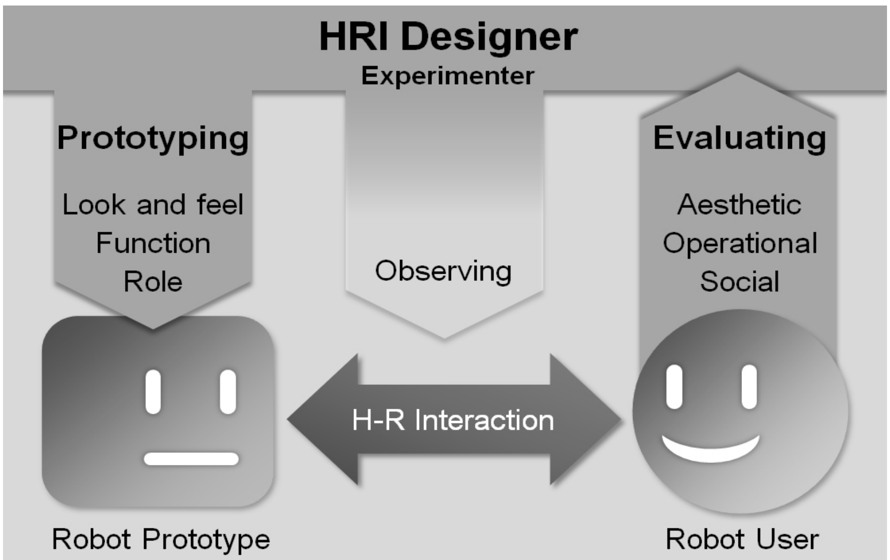


Fig. 5. Prototypes and elements for HRI research

These three prototypes and elements for HRI research can be schematized into an HRI research framework, as shown in Figure 5. This framework can not only facilitate the understanding of user-centered HRI, but also help novice researchers wishing to take a user-centered approach including designers, define HRI problems, and set research areas and methodologies.

4 User-Centered HRI Design Case Studies

This chapter is intended to help readers form a more concrete understanding of the proposed user-centered HRI design framework by introducing several research case studies that employ this framework. Case studies pertaining to the three categories of HRI design enumerated above—i.e. aesthetic contextuability, operational contextuability, and social contextuability—are presented below, with summaries of background, objectives, methodologies, and results.

4.1 Case Studies on Aesthetic Contextuability in HRI

4.1.1 Matching Proper Materials to the Role Images of Teaching Assistant Robots for Elementary School

The purpose of this case study was to extrapolate role images for a teaching assistant robot suited to elementary school students by focusing on one specific element of robot appearance: the materials used to manufacture a robot. This study presupposes that external materials constitute an important factor influencing the effective expression of a robot's role. The main topic of inquiry in this study, the materials forming a robot's exterior, falls under the heading of "aesthetic contextuability" in terms of user-centered HRI research; the significance of this study lies in its focus on "materials," an element of a robot's exterior that had heretofore been neglected by researchers. A robot's shape, color, and other basic elements of design have been addressed in previous studies as factors that can influence human-robot interaction (HRI). However, material—despite being an important element, along with shape and color, in the design development of robots—has been largely overlooked in HRI research thus far. Nonetheless, the materials that constitute a robot's exterior are closely connected to communication by touch as well as by sight; it is thus an element that deserves increasing attention as robots continue to be incorporated into the everyday living spaces of human beings.

The main purpose of this case study was to match materials with specific groups of adjectives relating to a robot's given role. Using a robotic platform with a simple structure and function, a wide range of materials including metal, glossy plastic, matte plastic, wood, fabric, and rubber were applied to the platform exterior to develop robot prototypes for testing (Figure 6). A prototype endowed with such form and function, known as a "look and feel" prototype, is a representative variety of prototype used in HRI research involving aesthetic contextuability. The tasks to be carried out by the test subjects were set through the robots' eye and head movement and through simple dialogue; while carrying out these tasks, the subjects were asked to evaluate the degree to which each robot's material accorded with its role image (adjective group) as a teaching assistant robot.

Fig. 6. Look and feel prototype used in the experiment

This case study was limited by its use of materials with colors unique to each material; further observation that separates touch-based cognitive responses to the material from visual cognitive responses to color is needed in future. Nonetheless, it clearly demonstrates that material can be an important factor in robot design for the effective expression of a robot's given role. Above all, this case study is valuable for its presentation of practical HRI research results that can be referenced by actual designers in the process of robot design development.

4.1.2 A Study on External Form Design Factors of Teaching Assistant Robots for Elementary School

The purpose of this case study was to identify the factors of a robot's physique that effectively express an image appropriate to a teaching assistant robot, and thereby propose guidelines for the external form design of teaching assistant robots. Like the study examined in 4.1.1, this study also has the ultimate goal of developing an optimal design for teaching assistant robots; instead of "materials," however, it focuses on a robot's "physique."

Prior to the experiment, images related to the role of teaching assistant robots were canvassed through textual research and a factor analysis, and eight factors of physique for teaching assistant robots were extracted from among the diverse factors that compose the human physique. Using these factors as variables, various 3D external form samples were created for testing.

These 3D robot models were projected in life size against the inner wall of the laboratory (Figure 7); they were viewed by children—the main user group for teaching assistant robots—who then filled out a questionnaire gauging their response to each model's physique. Whereas the study examined in 4.1.1 created physical look and feel prototypes, this study utilized 3D models that exist only in software form. However, it adopted an innovative new method for the use of "look and feel" prototypes: to maximize the subjects' sense of the 3D models' volume and presence, animation of each prototype in 360-degree rotation was created, and the models were presented in full life-size projection.

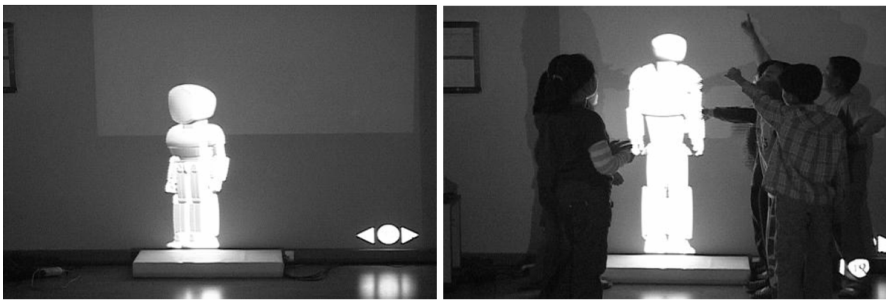


Fig. 7. 3D robot models presented in full life-size projection

The experiment results showed that the ratio of head length (vertical) to the body, height, ratio of chest circumference, and waist circumference were related to role images for teaching assistant robots. They further demonstrated that height and waist were particularly important, with waist circumference having a high correlation to all role images. To convey a gentle and kind image, for instance, the ratio of head length needed to be adjusted in relation to waist circumference.

Based on the results of analyzing the relationships among various factors of robot physique, this study proposed guidelines for designing the external form of teaching assistant robots. As such, this work does more than research the development of a particular type of robot, i.e. teaching assistant robots; rather, the methodology described in the study proposes new guidelines that can be used to research the development of new external forms for robots with different functions, uses, and user bases.

4.1.3 A Study on an Emotion Eliciting Algorithm and Facial Expression for Designing Intelligent Robots

Many humanoid robots that express emotion through facial expressions have been developed thus far. While it is easy to find HRI studies dealing with technical factors related to robot behavior, including facial expressions, it is relatively harder to find works, like the current case study, that rigorously address the perspective of the user who must interact with and use robots face-to-face. This case study can be classified as user-centered HRI research that focuses on the user's cognitive response to—and assessment of—robots.

The process of changing from one facial expression to another takes time, and the length of this process is affected by the type of emotion being expressed as well as by the strength of the external stimuli driving the expression change. The degree of "naturalness" conveyed by the robot's facial expression changes according to the dynamic between these factors, which in turn affects the "humanness" or "humanlike-ness" a human user feels when faced with a robot. This is the hypothesis informing the case study examined here (Figure 8). The study also falls under the heading of research on aesthetic contextuability, and, like the preceding examples, uses "look and feel" robot prototype s to test its experimental hypothesis.

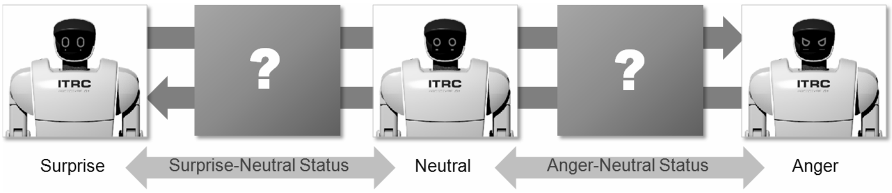


Fig. 8. The intermediary status of facial expression

As the technology for enabling robots to express emotion becomes increasingly sophisticated, the "humanness" of robots naturally increases as well. However, when the objective is to realize facial expressions under limited technological and budgetary conditions, enhancing the naturalness of the robot's emotional expression can be one way of increasing its humanness. This particular case study focuses on external form—specifically, elements relating to a robot's face and expressions—as a factor that increases humanness.

In this case study, four human expressions (anger, happiness, sadness, surprise) were selected for re-creation. A robot prototype capable of realizing these four expressions through alterations in the shape of its eyes and mouth was developed and used for experimentation. The main outcome of the study was the identification of the duration of facial expression change appropriate to each emotion, achieved by measuring the humanness of the robot as perceived by the subject in relation to the length of time required for the robot to change its facial expression. The relationship between the robot's emotional characteristics and facial expressions illuminated in the study suggests the need for more in-depth HRI research on a broader range of physical elements for emotional expression besides facial expressions. Above all, it supports the argument that the definition and classification of emotional models suited to robots, and of the corresponding tools and methods of expression, are needed to enable smooth human-robot communication and enhance a robot's capacity for task execution.

4.2 Case Studies on the Operational Contextuability of HRI

4.2.1 A Study on Humanoid Robot's Gesture Design

Among the many competitive advantages possessed by humanoid robots, the ability to gesture plays a particularly crucial role in enabling more intuitive and abundant communication during human-robot interaction. Gesture design also has a large impact on the determination of a robot's personality. However, there is as yet insufficient user-centered design research on what specific roles a robot's gestures serve during actual interaction with a human being, how gestures reflect a robot's personality, and what kinds of gestures are appropriate to the nature of various tasks. In this context, the current case study had three primary objectives: 1) to identify the impact of gestures in HRI, 2) to design a robot's personality through gesture design, and 3) to propose personality designs appropriate to the nature of various HRI tasks.

To achieve these goals, several experiments were conducted using an actual humanoid robot platform as a function prototype. The robot used in the experiments was a wheel-type humanoid robot named Amiet. The size, speed, and frequency of its upper-body movements—i.e. head, arm, and waist gestures—could be controlled; speech generation was enabled using TTS (text to speech) technology. Although Amiet was not developed specifically for this study, it was conducive to effectively designing the experiments in question, since it was a general humanoid type suitable for research relating to robotic gesture design. However, for optimized control of the various elements, human-robot dialogue scenarios and the corresponding speech and gestures were newly prototyped using the WOz technique (Figure 9).

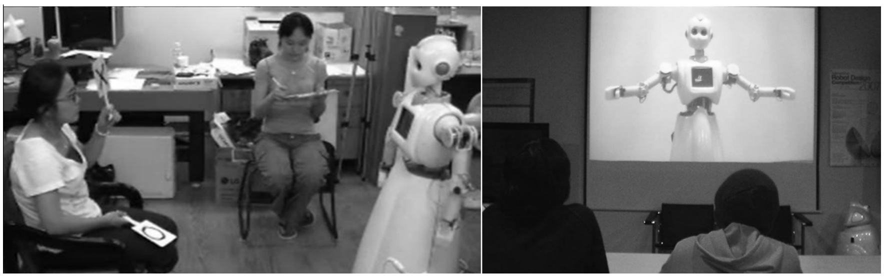


Fig. 9. Experiments using the humanoid robot platform

The discoveries made through the current case study are as follows: First, robotic gestures were found to have a stronger impact than speech when it came to effectively communicating meaning to human users; it was also found that a lack of consistency between gesture and speech left a negative impression on the user. Second, a robot's personality could be appropriately expressed, and perceived by the user, by controlling the size, speed, and frequency of its gestures. Third, depending on the type of personality derived from gesture design, there were different HRI tasks that were found to be contextually appropriate.

In conclusion, this case study effectively carried out experiments from the perspective of user-centered design by utilizing a function prototype to enable a fundamental understanding and application of humanoid gesture design, and provided for more intuitive usability through such experimentation. It further confirmed the operational contextuability of implementing gesture-based HRI design that is advantageous to the user's tasks.

4.2.2 Design of the Anthropomorphic Sound Feedback for Service Robot Malfunction

With the advancement of research fields related to service robots, human-robot interaction is taking on a more complicated aspect. As a result, users sometimes fail to understand System or structure of the robot and thus they are unable to respond appropriately to malfunction of robot. If an error occurs in a robot equipped with a powerful actuator, the user could even be placed in unexpected danger. Therefore, there is a strong need to look into the feedback design for the effective communication of such error situations (Lee et al., 2010). The majority of existing studies on robot feedback design address usability during normal operation; research on feedback for malfunction remains insufficient as yet.

The current case study aimed at applying anthropomorphic elements to feedback design for the malfunction of service robots, and at ascertaining the effects of such application on usability. It also attempted to discover how usability provided by anthropomorphic feedback differs according to the level of seriousness for errors. The cleaning robot, which has recently come into the spotlight as a commercialized everyday robot, was selected as the medium for achieving the goals of the experiment; to effectively control the various experimental factors, a function prototype possessing the external appearance of existing cleaning robots was created. All other elements were omitted from the prototype, besides the speaker, wheel motor and the wireless transmitting module and thus it could be controlled remotely just for the experimental purpose. For the scenario in which the cleaning robot was working normally, the sound of a vacuum cleaner played to induce the test subject to perceive that the robot was in cleaning operation. For malfunction situations, both anthropomorphic and non-anthropomorphic sound feedback was prototyped for two different levels of seriousness. Non-anthropomorphic feedback was simply mechanical sound while the anthropomorphic feedback did not have any linguistic functions and thus it sounded like higher animal such as dogs, dolphins and etc. Subjects evaluated the usability of feedback in each situation provided by the prepared scenarios in terms of its effectiveness, efficiency, and satisfaction. Video analysis of each subject's actions was also carried out (Figure 10).

Fig. 10. Experiments using a function prototype of cleaning robots

The findings of this study can be summarized as follows: 1) Anthropomorphic sound feedback was more effective to communicate general errors than its non-anthropomorphic counterpart. 2) When the level of seriousness was high, short and clear mechanical sound was more effective, since the situation demands urgency. 3) Therefore, it was clear that appropriate feedback design differed according to error type. As these findings illustrate, this study emphasized that response to error situations must be considered from the perspective of user-centered design, rather than through a technology-driven approach.

For credible incorporations with various robots, HRI design studies are responsible for developing the effective strategy to deal with a wide range of malfunctions of the robots. The conclusions derived from this case study provide inspiration regarding a new role for user-centered HRI design as a means of overcoming the negative aspects of HRI error that are bound to be experienced through the daily use of robots in the future.

4.2.3 Application of Unexpectedness to Behavioral Design for the Entertainment Robot

As products become increasingly intelligent, users have begun to expect pleasurable communication that goes beyond the effective execution of a product's given function. The most representative approach for enabling a product to provide pleasure to human users is interaction design for entertainment. Since robots can be seen as occupying the extreme end of the spectrum of intelligent products, they are naturally being required to provide entertainment to users. However, despite Sony's development of a variety of entertainment robots, including AIBO, QRIO, and Rolly, there are few instances of robot products that have been deemed successful. This is likely due to the fact that the design of robot behavior is severely limited as yet: the novelty effect of robot products quickly wears off as a result.

The current case study thus aimed at introducing unexpectedness to the behavioral design of entertainment robots. To this end, it specified three categories of unexpectedness and the methods used to express each type of unexpectedness, as well as the behavioral factors needed to generate a sense of fun. Based on these aspects, the study ascertained through experimentation how each type of expectation disconfirmation differently affects user perception.

Experiments were designed and carried out in two stages. In the first experiment, brief videos of robot behaviors expressing various types of unexpectedness were shown to the subjects, then they evaluated each behavior in terms of novelty, enjoyment, familiarity, performance, reliability, and satisfaction. This method, used to evaluate human-robot interaction scenarios, can be categorized as a video-based function prototype supported by study of Woods (2006), which showed that both video and live demonstrations produced virtually identical result values. The second experiment examined how the user's impression of, and satisfaction with, the robot differed according to the behavioral factors of unexpectedness. For effective experimentation, Robosapien V2, a toy robot currently on the market, was selected as a physical function prototype. Robosapien is a humanoid type robot measuring approximately 55cm; the actuators for its leg and arm joints can be controlled manually. The experiment consisted of direct interaction between subject and robot using the WOz technique in accordance with various specified scenarios; the subjects perceived and evaluated the robot's behavior (Figure 11).

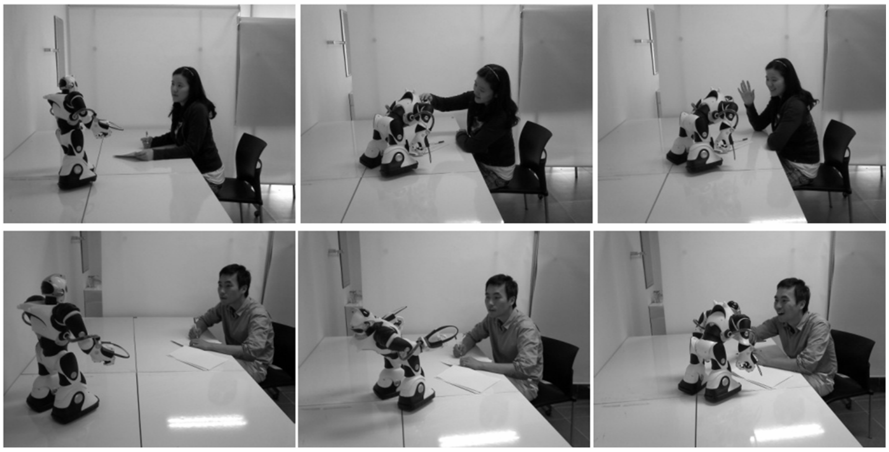


Fig. 11. Experiments using a humanoid toy

Based on the results of the study, explanations of user experience by type of unexpected behavior were presented, and behavioral design guidelines by target user were proposed. Such user-centered design knowledge points to an active role for HRI design that goes beyond simply providing improved usability through robot behavior, to aim ultimately at giving pleasure to human beings.

4.3 Studies on the Social Contextuability of HRI

4.3.1 Interaction Design of Teaching Assistant Robots Based on Reinforcement Theory

This case study is aimed at examining whether reinforcement theory, a brand of behaviorist learning theory, can be effectively applied to teaching assistant robots. Interaction systems were designed variously by robot type on the basis of reinforcement theory, and student's interactions with each robot type were observed. The process of student-robot interaction was recorded on video and in observation logs; these were used for a post-experiment analysis. After the experiments were completed, the changes wrought in student behavior by each robot type and the impressions received by students from the robots were analyzed.

Educational robots, which were the object of this case study, have been shown to convey a friendlier image than other media in assisting children's learning, and to be particularly effective in motivating young children (Han et al., 2006; Kanda et al., 2004). For this reason, they are garnering attention as a useful educational medium. However, to introduce robots to the educational scene, one must do more than simply use robots in education; it is necessary to apply various existing theories on education to the interaction design of robots, and to verify the effects by conducting studies involving children.

In the current case study, robots embodying three types of role models based on reinforcement theory—"Ching-chan-ee(rewarder)", providing positive reinforcement; "Um-bul-ee(punisher)", providing negative reinforcement, and "Sang-bul-ee(twofer)", providing both positive and negative reinforcement—were designed to examine changes in the interaction context of users. Test participants were required to interact with all three types of robots in a designated educational context, i.e. solving math problems, and were asked afterward to check their impression of and preferences for the various robots through interviews and questionnaires (Figure 12).

Fig. 12. Experiments for ascertaining the effects of role models for teaching assistant robots

This study demonstrated the need for teaching assistant robots in educational environments; it also proved that a student's impression of and preferences for teaching robots can vary according to the type of reinforcement used. Furthermore, regarding HRI research methodologies, it advocates the need not only for studies on user-robot interaction contexts in the field and in virtual usage environments, but also for those focusing on the relationship between user and robot.

4.3.2 Interaction Design of Teaching Assistant Robots Based on Reinforcement Theory

The purpose of this case study was to carry out a comparative analysis of the childcare-related behaviors of dual-income and single-income families, and to propose design guidelines for childcare robots tailored to each type of household. Unlike the studies examined in the preceding sections, the current study did not observe user-robot relationships in a laboratory environment, but rather observed actual behaviors in the field by visiting dual-income and single-income families selected for the study and analyzing journals kept by the users themselves.

This study canvassed the needs that arise in the interactions between mother and child either between user and product through in-depth interviews, home tours, and childcare journals. In the interviews, basic information relating to childcare, including the parents' personal information, the average amount expended on childcare, and educational facilities used, were obtained. In addition, questions on everyday childcare in dual-income and single-income families were posed, with clear distinction between weekday and weekend activities. After the interview, home tours were undertaken to observe and photograph the products and environments used for childcare within the home. In addition, the mothers of the various families were asked to keep a childcare journal over a one-week period, in which they stated their satisfaction level for each activity recorded and the product used at the time, along with the corresponding reasons.

The experiment results showed that needs relating to time and schedules, communication, and playmates arose more frequently in dual-income families, while needs relating to emotional communication were more frequent in single-income families. Needs relating to monitoring occurred with similar frequency in both types of families. Based on the needs thus observed, design guidelines for childcare robots capable of fulfilling such needs were devised. This study contributes to robot design research by showing that childcare robots need to be endowed with differing characteristics according to usage environment, family makeup, living pattern, and espoused values.

4.3.3 Focus Group Interview for Designing a Growing Robot

This case study aimed at observing long-term interaction between humans and their non-human counterparts for the purpose of designing a growing robot. To analyze interactions between a human and a physically changing, growing counterpart, focus group interview with people who keep puppies, plants, or products emulating plants was conducted.

Although most HRI studies focus on immediate responses caused by immediate perception, the novelty effect stemming from immediate interaction between human and robot drastically decreases once curiosity toward a new robot has dissipated. Therefore, sustained long-term interaction is needed to supplement this novelty effect and strengthen the social bond between a human user and a social robot.

In this case study, focus group interviews were conducted on subjects who keep a variety of growing animals and objects (dogs, cats, Flip Flops, etc.), and the interviews content was geared to gauge how the subjects' sense of closeness to their pet, plant, or plant substitute changed through time. This change, along with raising information, stimuli, and condition of the pet or plant, was mapped onto a graph as shown below (Figure 13).

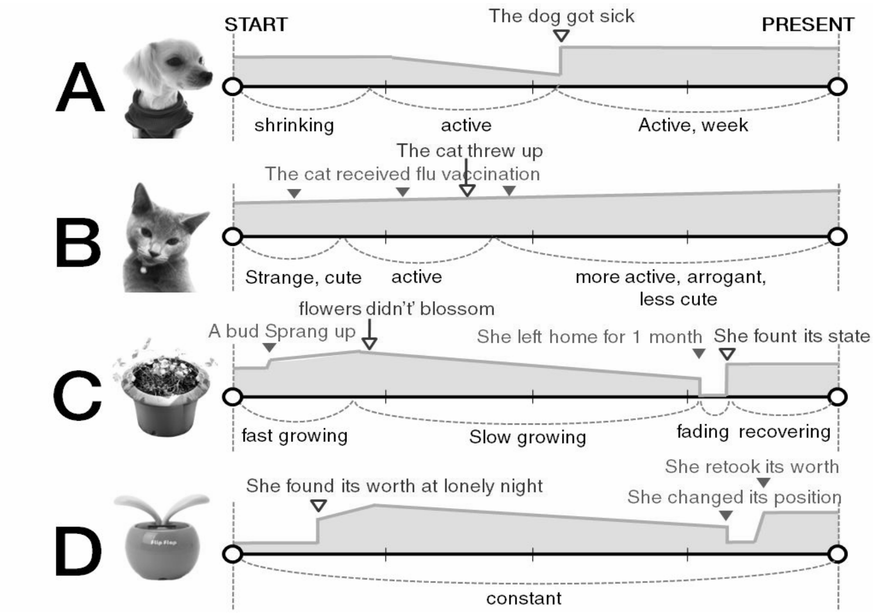


Fig. 13. Changes in the user's sense of closeness by object of relationship

The study arrived at the conclusion that people form special feelings for growing things or creatures, and see them as providing psychological stability, enjoyment, and recreational diversion. It also found that unexpected incidents and anticipated situations have a great impact on a user's actions and the sense of attachment he or she feels. Based on these results, this study can be said to have significance as basic research for the development of long-term interaction design between human users and growing robots.

5 Conclusion

Until now, robots have existed primarily as objects of experimentation within laboratory settings. However, robotics research is ultimately aimed at enriching the quality of life for human beings. Therefore, the need for realistic user-centered HRI design research is being raised as a means to create robots capable of fulfilling roles as helpers and partners for humans in everyday life. This chapter examined the definition, function, and methodology of user-centered design research, which supplements the robot-centered technological development that heretofore formed the main method of research on human-robot interaction (HRI). User-centered HRI design research focuses on the user's senses, behaviors, and values in looking at design problems; this focus can be divided into the three elements of aesthetic contextuability, operational contextuability, and social contextuability. Accordingly, this type of research uses intuitive approaches that employ "look like," "function," and "role" prototypes, which users can actually feel, communicate with, and judge. For each field of design addressed in user-centered HRI research, this chapter introduced the purpose, method, and results of representative case studies. Looking at recent trends in HRI research, it is possible to discern a transition from studies dealing with aspects of immediate sensory experience to those focusing on actions and behaviors; in the future, increasing emphasis will be placed on issues of value. None of the case studies introduced above are confined exclusively to the realm of aesthetic, operational, or social contextuability. This is because robots, by nature, perform interactions that are always accompanied by all three aspects. In the end, to design robots that can coexist naturally and usefully with humans in everyday life, a harmonious balance must be struck among the three elements of HRI research identified in this study.

The field of HRI research is so profound, extensive, and complicated that a human being engaging in such research may be compared to a creator making a creature, or robot, in his or her own image. For this reason, the boundaries between academic disciplines must be increasingly lowered and researchers in engineering, design, and the humanities must come together to ponder the many problems that remain.

References

Bartneck, C., Forlizzi, J.: A design-centered framework for social Human-robot interaction. In: Proceedings of the 13th IEEE International Workshop on Robot and Human Interactive Communication, Piscataway, USA, pp. 31–33 (2004)

Bartneck, C.: Rapid prototyping for interactive robots, In: Proceedings of the 8th Conference on Intelligent Autonomous Systems(IAS-8), [CD ROM], Amsterdem, Netherland (2004)

Breazeal, C.: Social interactions in HRI: The robot view. IEEE Transactions on Systems, Man and Cybernetics, Part C: Applications and Reviews 34(2), 181–186 (2004)

Buchanan, R.: Rhetoric, humanism, and design. In: Buchanan, R., Margolin, V. (eds.) Discovering design: Explorations in design studies, pp. 23–66. University of Chicago Press, Chicago (1995)

Buchanan, R.: Design research and the new learning. Design Issues 17(4), 23–30 (2001)

Card, S.K., Moran, T.P., Newell, A.: The Psychology of Human-Computer Interaction. CRC Press, Boca Raton (1983)

Cross, N.: Design/Science/Research: Developing a discipline. In: Proceedings of the 5th Asian Design Conference, [CD ROM], Seoul, Korea (2001a)

Cross, N.: Designerly ways of knowing: Design discipline versus design science. Design Issues 17(3), 49–55 (2001b)

Dautenhahn, K.: Robots as social actors: AuRoRA and the case of autism. In: Proceedings of the 3rd Conference on Cognitive Technology, Los Alamitos, USA, pp. 359–374 (1999)

Fong, T., Nourbakhsh, I., Dautenhahn, K.: A Survey of socially interactive robots. Robotics and Autonomous Systems 42(3-4), 143–166 (2003)

Forlizzi, J.: How robotic products become social products: An ethnographic study of robotic products in the home. In: Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction, New York, USA, pp. 129–136 (2007a)

Forlizzi, J.: The product ecology: Understanding social product use and supporting design culture. International Journal of Design 2(1), 11–20 (2007b)

Gould, J.D., Conti, J., Hovanyecz, T.: Composing letters with a simulated listening typewriter. Communications of the ACM 26(4), 295–308 (1983)

Green, S.W., Jordan, W.P.: Pleasure with Products: Beyond Usability. CRC Press, Boca Raton (2002)

Han, J., Kim, D.: Field trial on robots as teaching assistants and peer tutors for children. In: Proceeding of the Asia Pacific International Symposium on Information Technology, [CD ROM], Hangzhou, China (2006)

Hassenzahl, M.: The effects of perceived hedonic quality on product appealingness. International Journal of Human-Computer Interaction 13(4), 479–497 (2002)

Kanda, T., Hirano, T., Eaton, D., Ishiguro, H.: Interactive robots as social partners and peer tutors for children: A field trial. Human-Computer Interaction 19(1-2), 61–84 (2004)

Kim, J., Kim, M.: A fundamental study design-centered HRI research framework. In: Proceedings of 2005 Bi-annual Design Conference of Korea Society of Design Science, Seoul, Korea, pp. 68–69 (2005)

Lee, M., Kiesler, S., Forlizzi, J., Srinivasa, S., Rybski, P.: Gracefully mitigating breakdowns in robotic services. In: Proceedings of Human Robot Interaction 2010, [CD ROM], Osaka, Japan (2010)

Lobach, B.: In: Lee, B. (ed.) Industrial design, Jo-hyung Kyou-youk, Seoul (2000)

Mugge, R., Schoormans, J.P.L., Schifferstein, H.N.J.: Product attachment: Design strategies to stimulate the emotional bonding to products. In: Schifferstein, H.N.J., Hekkert, P. (eds.) Product experience, pp. 425–440. Elsevier, New York (2008)

Norman, D.A.: The psychology of everyday things. Basic Books, New York (1998)

Norman, D.A.: Emotional design. Basic Books, New York (2004)

Oh, K., Kim, J., Kim, M.: Development of humanoid robot design process-focused on the concurrent engineering based humanoid robot design. In: Proceedings of IASDR 2005 Conference, [CD ROM], Yunlin, Taiwan (2005)

Oh, K., Kim, M.: The HRI experiment framework for designers. In: Proceedings of IASDR 2007 Conference, [CD ROM], Kolon, Hong Kong (2007)

Schon, D.A.: The reflective practitioner. Basic Books, New York (1983)

Simon, H.A.: The science of the artificial. The MIT Press, Cambridge (1969)

Swann, C.: Action research and the practice of design. Design Issues 18(1), 53–54 (2002)

Wood, S.N., Walters, M.L., Koay, K.L., Dautenhahn, K.: Methodological issues in HRI: A comparison of live and video-based methods in robot to human approach direction trials. In: Proceedings of Human Robot Interaction Conference 2006, [CD ROM], Utah, USA (2006)

Yang, M.C.: An examination of prototyping and design outcome. In: Proceedings of Design Engineering Technology Conferences and Computers in Information and Engineering Conference 2004, [CD ROM], Salt Lake City, USA (2004)

What Is Mixed Reality, Anyway? Considering the Boundaries of Mixed Reality in the Context of Robots

J. Young[1,2], E. Sharlin[1], and T. Igarashi[2,3]

[1] The University of Calgary, Canada
[2] The University of Tokyo, Japan
[3] JST ERATO, Japan

Abstract. Mixed reality, as an approach in human-computer interaction, is often implicitly tied to particular implementation techniques (e.g., see-through device) and modalities (e.g., visual, graphical displays). In this paper we attempt to clarify the definition of mixed reality as a more abstract concept of combining the real and virtual worlds – that is, mixed reality is not a given technology but a concept that considers how the virtual and real worlds can be combined. Further, we use this discussion to posit robots as mixed-reality devices, and present a set of implications and questions for what this implies for mixed-reality interaction with robots.

Keywords: Human-robot interaction, mixed reality, human-computer interaction.

1 Introduction

Mixed reality is a popular technique in human-computer interaction for combining virtual and real-world elements, and has recently been a common technique for human-robot interaction. Despite this popular usage, however, we argue that the meaning of "mixed reality" itself is still vague. We see this as a challenge, as there is a great deal to be gained from mixed reality, and a clear definition is crucial to enable researchers to focus on what mixed reality offers for interaction design.

In this paper, we attempt to clarify the meaning of mixed reality interaction, and follow by relating our discussion explicitly to human-robot interaction. In short, we propose that mixed reality is a concept that focuses on how the virtual and real worlds can be combined, and is not tied to any particular technology. Based on our definition we posit that robots themselves are inherently mixed-reality devices, and demonstrate how this perspective can be useful for considering how robots, when viewed by a person, integrate their real-world manifestation with their virtual existence. Further, we outline how viewing robots as mixed reality interfaces poses considerations that are unique to robots and the people that interact with them, and raises questions for future research in both mixed reality and human-robot interaction.

X. Wang (Ed.): Mixed Reality and Human-Robot Interaction, ISCA 47, pp. 1–11.
springerlink.com

2 Considering Boundaries in Mixed Reality

> *Mixed Reality* – “**Mixed reality** refers to the merging of real and virtual worlds to produce new environments and visualisations where physical and digital objects co-exist and interact in real time.”[1]

The above definition nicely wraps the very essence of what mixed reality is into a simple statement – mixed reality merges physical and digital worlds. In contrast to this idea-based perspective, today mixed reality is often seen as a technical implementation method or collection of technologies. In this section, we attempt to pull the idea of mixed reality away from particular technologies and back to its abstract and quite powerful general essence, and highlight how this exposes some very fundamental, and surprisingly difficult, questions about what exactly mixed reality is. In particular, we show how robots, and their inherent properties, explicitly highlight some of these questions.

We start our discussion by presenting research we conducted (Young and Sharlin, 2006) following a simple research question: given mixed reality as an approach to interaction, and, robots, we asked ourselves: “if we completely ignore implementation details and technology challenges, then what types of interactions does mixed reality, as a concept, enable us to do with robots?” In doing this, we forced ourselves to focus on what mixed reality offers in terms of interaction possibilities, rather than what we can do with a given implementation technology, e.g., a see-through display device, or the ARToolkit[2] tracking library. We formalized this exploration into a general idea for mapping such an interaction space, and presented exemplary techniques (Young and Sharlin, 2006) – we present the core of this work below, where the techniques serve as interaction examples to be used throughout this paper.

2.1 The Mixed Reality Integrated Environment (MRIE)

Provided that technical and practical boundaries are addressed, the entire three-dimensional, multi-modal real world can be leveraged by mixed reality for integrating virtual information. One could imagine a parallel digital, virtual world superimposed on the real world, where digital content, information, graphics, sounds, and so forth, can be integrated at any place and at any time, in any fashion. We called such an environment the “mixed-reality integrated environment”, or the MRIE (pronounced “merry”) (Young and Sharlin, 2006), and present it as a conceptual tool for exploring how robots and people can interact using mixed reality. Specifically, we used the MRIE as a technology-independent concept to develop a taxonomy that maps mixed-reality interaction possibilities (Young and Sharlin, 2006), and used this taxonomy to devise specific interaction techniques. For our current discussion, we quickly revisit two of the interaction techniques we proposed in our MRIE work: bubblegrams and thought crumbs (Young and Sharlin, 2006).

---

[1] http://en.wikipedia.org/wiki/Mixed_reality, retrieved 11/11/09.
[2] http://www.hitl.washington.edu/artoolkit/

Bubblegrams – based on comic-style thought and speech bubbles, bubblegrams are overlayed onto a physical interaction scene, floating next to the robot that generated it. *Bubblegrams* can be used by the robot to show information to a person, and can perhaps be interactive, allowing a person to interact with elements within the bubble (Figure 1).

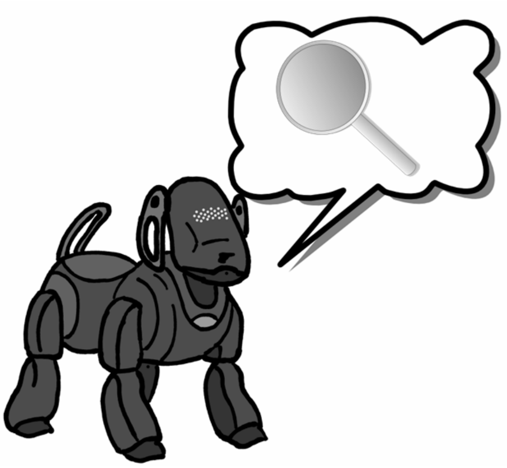


Fig. 1. Bubblegrams

Thought Crumbs – inspired by breadcrumbs from the Brothers Grimm's *Hansel and Gretel*[3], *thought crumbs* are bits of digital information that are attached to a physical, real-world location (Figure 2). A robot can use these to represent thoughts or observations, or a person could also leave these for a robot to use. These can also perhaps be interactive, offering dynamic digital information, or enabling a person or robot to modify the though crumb.

Fig. 2. Thought crumbs, in this case a robot leaves behind a note that a person can see, modify, or interact with later

[3] http://en.wikipedia.org/wiki/Hansel_and_Gretel

2.2 Basic Implementation

Our original bubblegrams implementation (Figure 3) uses either a head-mounted or a tablet see-through display, where the head mounted display setting was used for viewing only, and interaction was only possible through the tablet setting. Using a vision algorithm, the location of the robot is identified in the scene and the bubble is drawn on the display beside the robot. A person can interact with the bubble using a pen on the tablet PC (Young et al., 2005).

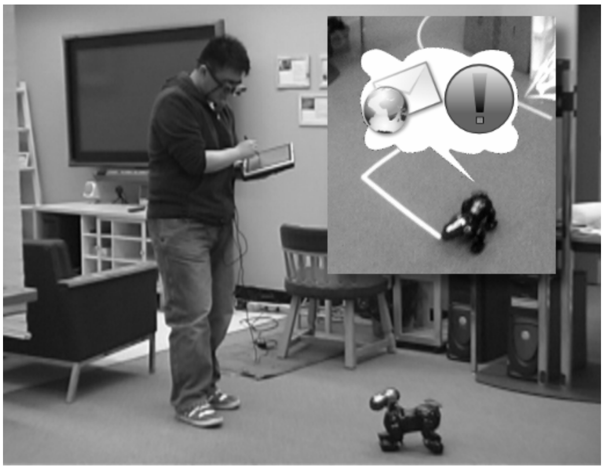


Fig. 3. Bubblegrams see-through device implementation

Few would argue that this is a mixed-reality system, as it fits a very common mixed-reality implementation mould – see-through display with computer graphics superimposed over real-world objects. However, consider the case where an interface designer does not want to use a bulky hand-held display and opts to replace the graphical bubbles with, perhaps, a display attached to the robot. This display would show the exact same information as in the prior interface but would not require the person to carry any actual equipment – is this still mixed reality?

Perhaps the designer later decides to replace the display with a series of pop-out cardboard pieces, with a clever set of retractable cut-outs and props – possibly mounted on springs to add animation effects. While we concede that there are important differences with this approach, such as a greatly-reduced level of flexibility, this display still represents digital, virtual information and superimposes it in the real world in much the same way (conceptually) as the previous method – is this still mixed reality?

The thought crumbs implementation (Figure 4) uses RFID tags for messages, where the physical tag itself denotes the location of the message, and the message information is stored within the tag. The tags also have human-readable outward appearances, and are supplemented with infrared lights so the robot can locate the tags from a distance (Marquardt et al., 2009). In a similar effort, *Magic Cards*

(Zhao et al., 2009), paper tags are used by both the person and the robot. A robot can leave representations of digital states or information at meaningful real-world locations as paper printouts, and can read cards left by people, enabling a person to interact with the robot's virtual state through working with physical cards.

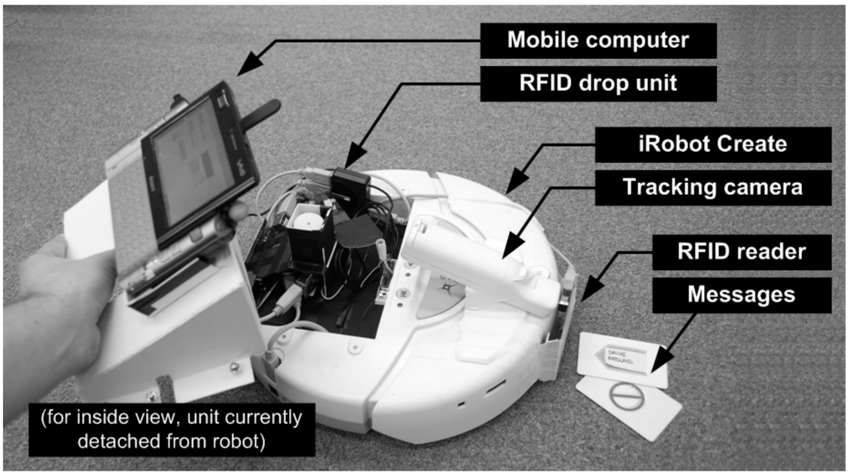


Fig. 4. RFID Thought Crumbs implementation

Our original thought crumbs discussion (Section 2.1) introduced it as a mixed-reality interaction technique, and in both the implementations shown here virtual information (pending robot commands, system state, robot feedback, etc) is integrated into the physical world through their manifestations. Overall the core concept of the interaction is the same as the original idea, but are these implementations, without any superimposed visual graphics, mixed reality?

The above discussion highlights how easy it is to draw lines on what kinds of interaction or interfaces count as mixed reality, based solely on the implementation technology. We fear that this can serve as a limiting factor when exploring mixed-reality techniques for interaction with robots, and argue that mixed reality should not be limited to or limited by any particular technology, implementation technique, or even modality (graphics, audio, etc). We see the concept of mixed reality itself as a very powerful approach to interaction, one that can serve as motivation for a plethora of interaction techniques and possibilities far beyond what is possible by the current technical state-of-the-art.

3 Defining Mixed Reality

Should mixed reality be viewed as an interaction device or mechanism, similar to a WiiMote or a tabletop? Or as an implementation tool such as C# or ARToolkit[4]

[4] http://www.hitl.washington.edu/artoolkit/

that enables the superimposing of computer graphics via a display device onto the real world? Is mixed reality limited to particular modalities, such as graphics, or, can it include other modalities such as sound or haptic interaction? Or, is mixed reality a more-general approach, such as ubiquitous computing, that spans particular implementations, tools, or modalities?

The common-use definition of mixed reality is difficult to pinpoint, but we believe that it is summed up by our earlier quote (Section 2). Note that this definition itself reveals a muddled stance. On the one hand it clearly describes the general idea of merging of real and virtual worlds. On the other hand, it explicitly focuses the definition toward the modality, “visualizations”. This limits and shapes the perspective offered by the definition, where we argue that mixed reality transcends the modalities.

3.1 Milgram and Kishino

In 1994 Milgram and Kishino presented what is considered to be a seminal discussion of mixed reality (Milgram and Kishino, 1994). This paper’s self-proclaimed primary contribution is a taxonomy of graphical, visual displays, and as such the tone of the paper surrounds mixing graphical and real-world environments.

On closer inspection, however, the theoretical discussion of the paper, including the well-known “virtuality continuum”, leaves the visual focus behind and is careful to abstract to the more general case. They say mixed reality is combining “real” objects, those that have an “actual objective existence”, with “virtual” objects, those objects which exist “in effect, but not formally or actually.” Further, the authors directly state that their focus on visual displays is largely related to the current state of technology, and outline that, as technology allows, mixed reality will include, for example, “auditory displays” and “haptic displays” (Milgram and Kishino, 1994). Below we attempt to relate this broader view of Milgram and Kishino’s model to current state of the art in tangible, physical and robotic interaction.

3.2 Mixed Reality and Tangible-User Interfaces

Much of the work in tangible computing revolves around the observation that we, as computer users, are simultaneously living in two realms: the physical one and the virtual one. Tangible user interfaces, then, are devices that are designed to “augment the real physical world by coupling this digital information to everyday physical objects and environments” (Ishii and Ullmer, 1997).

A modality-independent, general definition of mixed reality can be applied to a number of interaction approaches, with physical/tangible interaction being a straightforward extension. Strong parallels can be found between the motivation and meaning behind tangible-user interfaces and the general mixed reality approach of combining the virtual and the physical. Particularly if we discard the technology used or the communication modality (graphics, haptics, aural, etc) it becomes clear that both approaches are similarly attempting to find ways to combine the virtual and the real.

With this we do not mean to lessen tangibles or to imply any debasement as a research area, but to rather bring tangibles, and common understanding of mixed reality, under the same general theoretical foundation. We hope that this unification can help provide focus to the real challenges (and real contributions) that are being faced by these fields. Particularly, we are interested on focusing on interaction, more specifically human-robot interaction, and not any particular implementation tools or technologies.

3.3 Revisiting the Meaning of Mixed Reality Interaction

We see mixed reality as the concept of meshing the virtual and physical worlds into one interaction space. If we accept this definition, then there is an immediate problem of scope. For example, would not a mouse, as it couples physical input to virtual cursor state, or even a monitor, which gives a real view (via the photons it emits) of a virtual space, be a mixed reality device? This wide scope raises the question of how this broad definition can be useful or even desirable.

Mixed reality, as a concept, helps to push thinking toward the combination between the virtual and the real. It is useful as a sensitizing concept, or as tool to explicitly focus on the point of meshing. While the mouse is an amazingly successful interface in general, mixed reality highlights the mouse's limitations to mesh the virtual and the real – the link is unidirectional (no inherent physical feedback from the virtual world) and limited to the mouse's two-dimensional relative movements.

As another example, the *Magic Cards* interface described above (Zhao et al., 2009) uses physical print-out cards as a representation of a robot command or feedback message. Mixed reality points out that the paper (and printer) is the medium and sole contact point for bridging the virtual and the physical, and pushes us to consider how real information (e.g., location, real-world tasks) and virtual information (e.g., robot commands, robot feedback) can be linked through this interface. The same analysis applies for the thought crumb implementation presented earlier, (Marquardt et al., 2009), where RFID tags couple digital information with a particular real-world location (denoted by the location of the tag itself).

While this wide scope may sometimes make it difficult to draw lines on what mixed reality constitutes, thinking of interaction as mixed reality is useful as a tool that explicitly pushes us to consider the mapping between virtual objects, views, or states and the real-world and physical manifestations.

3.4 What Mixed Reality Provides

The idea of mixed reality as we present it provides only a simple, overarching perspective on interaction and is itself a very limited tool for examining, describing, and exploring interaction. That is, our approach does not supplant existing frameworks, categorizations, or interface design practices. Rather, mixed reality is a point of view from which existing tools can be applied.

For example, we do not consider how to approach interaction or interface design or evaluation, in either the real or virtual worlds. Existing design philosophies, heuristics, and so forth, still apply; mixed reality points toward the meshing point between the virtual and the real.

Further, we do not discuss how such a meshing point could be considered, targeted, mapped, and so forth, as this is already an active area of work in HCI. For example, mixed-reality work like Milgram and Kishino's virtuality continuum (Milgram and Kishino, 1994), tangible computing work such as Sharlin et al.'s consideration of input-/output-space coupling (Sharlin et al., 2004), or even by concepts such as Dourish's "embodied interaction," where the meaning of interaction (and how interaction itself develops meaning) is considered within the tangible and social real-world context (Dourish, 2001). Our approach on mixed reality shows how work such as this can be brought together under a common conceptual foundation.

To summarize, we view mixed reality not as a given technology or technique but as an interaction concept that considers how the virtual and real worlds can be combined into a unified interaction space. Therefore, rather than trying to decide if an interface incorporates mixed reality or not, we recommend that mixed reality itself be used as a tool to help directly consider the convergence points where the virtual and real meet.

4 Robots and Mixed Reality

So far, most of the mixed-reality discussion in this paper could be applied without any particular concern for robots. In this section, we outline how robots bring unique considerations to the table for mixed reality.

4.1 Agency

Robots are unique entities in that they have clearly-defined physical, real-world manifestations, can perform physical actions, and can act with some level of autonomy – this sets them apart from other technologies such as the PC (Norman, 2004). These real-world actions can easily be construed as life-like, and people have a tendency to treat robots similar to living entities, for example by anthropomorphizing, and give robots names, genders, and ascribe personalities (Forlizzi and DiSalvo, 2006; Sung et al., 2007). As part of this, people have been found to readily attribute intentionality and agency to robots and their actions. While people attribute agency to, e.g., video game characters and movies (Reeves and Nass, 1996), robots' real-world abilities and presence give them a very distinct, physically-embedded sense of agency that sets robots apart from other technologies.

In some ways, then, interacting with a robot has similarities with interacting with an animal or a person (Young et al., 2008a). The robot itself is seen as an independent, capable entity, and there is a sense of ownership and responsibility that ties the interactions with the robot, and the results of the interactions, back to the robot "entity" itself.

4.2 Mixed-Reality Entities

Robots are mixed reality entities, simultaneously virtual and real. They are virtual in that they are, essentially, a computer with virtual states, abilities, calculations, and a wide range of data in any number of formats. They are real entities in their physical manifestation, where they can interact with the world through this manifestation, both manipulating the world (output) and sensing it (input). As such, we argue that robots are, by their very nature, mixed-reality entities, as a large part of what makes them a robot is how they span the virtual and real worlds – the robot *itself* is a direct coupling of the virtual and the real.

Robots, as mixed reality interfaces, have a very explicit coupling between their virtual and real components. Due to agency, the various (virtual and real) components of the robot are directly attributed to (perhaps owned by) the individual, underlying conceptual agent (robot). The agent itself is directly tied to both the physical and virtual manifestations. This series of connections, supported by agency, means that interacting with robots is fundamentally different from interacting with interfaces that do not have agency; we attribute our interactions with the virtual and physical components directly to the underlying agent.

5 Discussion

We have argued for a wide view on mixed reality, and that robots themselves are inherently mixed-reality devices. What exactly this implies for human-robot interaction with mixed reality is not yet clear, and this is an important area for future consideration. In this section, we outline a few particular questions and challenges raised by this framing that we feel are important to consider.

Ownership and Boundaries – the consideration that robots have a strong sense of agency, coupled with their explicit, physical manifestation, raises questions of ownership and boundaries. For one, robots can (through technical means) claim ownership and enforce interaction constraints on mixed reality elements (Young and Sharlin, 2006). However, does this idea of robot / non-robot / human ownership of mixed-reality entities and items make sense to people in practice? If so, how can such ownership be mitigated and organized? Does this relate to concepts of virtual ownership we are familiar with, such as file permissions, private blogs, or even online finances? Similarly, are their implied boundaries in both the physical world and virtual worlds surrounding the robot as they may surround a living entity, such that, even without explicit ownership, people are careful about interacting in the robots *personal* space? Finally, is there a conceptual difference between the robot's mixed-reality *thoughts* (observations, etc), and ones drawn from the larger virtual world, such as the internet?

Agency – robots are not the only mixed-reality entities to have agency, with a simple example being animated, graphical mixed-reality characters. In this paper we argue that robotic agency is unique for various reasons, but this stance needs to be investigated further: is robot agency different enough from animated mixed-reality characters to merit special consideration? We are currently exploring this through comparing an animated system (Young et al., 2008b) to a very similar robotic system (Young et al., 2009). Further, if this is the case, what does this difference mean for the design of and interaction with mixed-reality interfaces? Following, the above personal-space concerns explicitly apply to the physical body (and perhaps any virtual manifestation) of the robot – do people have reservations about meddling with the robot itself as they may have for animals or people?

Interaction – if robots are simultaneously virtual and real entities, then what does this mean for mapping interaction with the robot? For example, is there a difference between on-robot-body techniques, such as embedded displays, direct haptic interaction (e.g., handshake), or robot sounds, and off-body techniques, such as projected displays, or thought crumbs left behind? How can people interact with these different types of interfaces?

6 Conclusion

In this chapter, we made the argument for moving the ideas of mixed reality away from the constraints of any particular implementation method or technique, or interaction modality – mixed reality is simply the mixing of the virtual and the real. Robots, then, fall under this wide perspective as inherently mixed reality devices that simultaneously exist in both realms. This perspective enables us to focus directly on the points of meshing between the virtual and the real, and the interface challenges and decisions related to making this meshing happen the way we want it to.

There are still many questions and challenges to be answered surrounding this outlook. Viewing robots as mixed reality devices does not change what we can do with robots, but it does provide us with a perspective that highlights how a robot exists both in the virtual and real realms, and, we hope, encourages us to consider what this means for interaction.

References

Dourish, P.: Where the Action Is: The Foundation of Embodied Interaction. The MIT Press, Cambridge (2001)

Forlizzi, J., DiSalvo, C.: Service robots in the domestic environment: a study of the roomba vacuum in the home. In: Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction, HRI 2006, Salt Lake City, USA, March 2-4, 2006, pp. 258–266. ACM Press, New York (2006)

Ishii, H., Ullmer, B.: Tangible bits: Towards seamless interfaces between people, bits and atoms. In: ACM Conference on Human Factors in Computing Sysems, CHI 1997, Atlanta, GA, March 22-27, pp. 234–241. ACM Press, New York (1997)

Marquardt, N., Young, J.E., Sharlin, E., Greenberg, S.: Situated messages for asynchronous human-robot interaction. In: Adjunct Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (Late-Breaking Abstracts), HRI 2009, San Diego, US, March 11-13. ACM Press, New York (2009)

Milgram, P., Kishino, F.: A Taxonomy of Mixed Reality Visual Displays. IEICE Transactions on Information and Systems E77-D(12), 1321–1329 (1994)

Norman, D.: Emotional design: why we love (or hate) everyday things. Basic Books, New York (2004)

Reeves, B., Nass, C.: The Media Equation: How people treat computers, television, and new media like real people and places. CSLI Publications, Center for the Study of Language and Information Leland Standford Junior University, Cambridge, UK (1996) (first paperback edition)

Sharlin, E., Watson, B., Kitamura, Y., Kishino, F., Itoh, Y.: On tangible user interfaces, humans and spatiality. Personal and Ubiquitous Computing 8(5), 338–346 (2004)

Sung, J.-Y., Guo, L., Grinter, R.E., Christensen, H.I.: "My roomba is rambo": Intimate home appliances. In: Krumm, J., Abowd, G.D., Seneviratne, A., Strang, T. (eds.) UbiComp 2007. LNCS, vol. 4717, pp. 145–162. Springer, Heidelberg (2007)

Young, J.E., Hawkins, R., Sharlin, E., Igarashi, T.: Toward acceptable domestic robots: Lessons learned from social psychology. Int. J. Social Robotics 1(1) (2008a)

Young, J.E., Igarashi, T., Sharlin, E.: Puppet master: Designing reactive character behavior by demonstration. In: ACM SIGGRAPH/EG SCA 2008, pp. 183–191. EG Press, EG Association, Germany (2008b)

Young, J.E., Sakamoto, D., Igarashi, T., Sharlin, E.: Puppet master: A technique for defining the actions of interactive agents by demonstration. In: Proc. HAI Symposium 2009, Tokyo (December 2009) (in Japanese) (presentation)

Young, J.E., Sharlin, E.: Sharing Spaces with Robots: an Integrated Environment for Human-Robot Interaction. In: Proceedings of the 1st International Symposium on Intelligent Environments, ISIE 2006, Cambridge, UK, April 5-7, pp. 103–110. Microsoft Research Ltd. Press, Cambridge (2006)

Young, J.E., Sharlin, E., Boyd, J.E.: Implementing bubblegrams: The use of haar-like features for human-robot interaction. In: Proceedings of the Second IEEE Conference on Automation Science and Engineering, CASE 2006, Shanghai, China, October 8-10, pp. 298–303. IEEE Computer Society Press, Los Alamitos (2005)

Zhao, S., Nakamura, K., Ishii, K., Igarashi, T.: Magic cards: A paper tag interface for implicit robot control. In: Proceedings of the ACM Conference on Human Factors in Computing Systems, CHI 2009, Boston, US, April 4-9, ACM Press, New York (2009)

Author Biographies

Krzysztof Arent received the M.Sc. degree in control engineering from the Wrocław University of Technology, Poland, in 1991 and a Ph.D. degree in applied mathematics from the University of Twente, The Netherlands, in 1995. He currently works as an assistant professor at the Institute of Computer Engineering, Control and Robotics, Department of Electronics, Wrocław University of Technology. His current research interests include the dynamics and control of robotic systems, adaptive control and social robots.

Jeong-gun Choi is a PhD candidate in the Department of Industrial Design at KAIST. He has been a researcher in Product and Environmental System Design Laboratory in KAIST since 2005. His current research is concerned with anthropomorphism and Human Robot Interaction. He has worked on robot design research projects with several domestic research institutions or companies (eg. ETRI, KT, Hanwool Robotics, ITRC-IRRC). He is also a multi-disciplinary design practitioner with experience spanning product, communication and environmental design.

J. Edward Colgate received the Ph.D. degree in mechanical engineering in 1988 from M.I.T. He subsequently joined Northwestern University, where he is currently a professor in the Department of Mechanical Engineering. Dr. Colgate's principal research interest is human-robot interaction. He has worked extensively in the areas of haptic interface and teleoperation, and he (with Michael A. Peshkin), is the inventor of a class of collaborative robots known as "cobots." He has served as an associate editor of the Journal of Dynamic Systems, Measurement and Control and the IEEE Transactions on Robotics, and he is the founding Editor-in-Chief of the IEEE Transactions on Haptics. He is also a cofounder of Stanley Cobotics and of Kinea Design. Dr. Colgate is currently the co-Director (with Don Norman) of the Segal Design Institute at Northwestern University.

Kerstin Dautenhahn is Professor in the School of Computer Science at the University of Hertfordshire, where she coordinates the Adaptive Systems Research Group. She received her PhD degree from the Biological Cybernetics Department, University of Bielefeld, Bielefeld, Germany, in 1993. She has pioneered research in social robotics, human–robot interaction and assistive technology, and has published more than 200 research articles. She has edited several books and frequently gives invited keynote lectures at international meetings. She has been Principal Investigator of her research team in several European projects. She is Editor-in-Chief of the journal Interaction Studies: Social Behavior and Communication in Biological and Artificial Systems.

Brian P. DeJong received the Ph.D. degree in mechanical engineering from Northwestern University in 2007. His doctoral research focused on haptic cyclic robots for lower limb exercise. Since then, he has been a professor of mechanical engineering in the School of Engineering and Technology at Central Michigan University. His primary research interest is in using auditory occupancy grids with a mobile robot, although he has concurrent research in robotics, teleoperation, human-robot interfaces, lower-limb exercise robots, and haptics.

Mohammad Firoozabadi received the B.Sc. Degree in electronics engineering from University of Tabriz, in 1987, the M.Sc. degree in electronics engineering from Amir-Kabir University of Technology, Tehran, in 1991, and the Ph.D. Degree in electrical engineering (biomedical engineering) from Tarbiat Modares University, Tehran, Iran, in 1997.

His research and teaching interests include theory and application of Human-Machine Interaction, bio-electromagnetics, bioelectric phenomena and electrophysiology, bio-instrumentation, and biological signal processing. Prof. Firoozabadi currently serves as a Professor of Biomedical Engineering, Deputy Dean of Medical Sciences Faculty at Tarbiat Modares University, and President of Iranian Society of Biomedical Engineering. Mohammad is the founder of Center of Advanced Researches and Technologies for Empowering and Mind Engineering (CARTEME) at Science and Research Branch, Islamic Azad University.

He has published over 220 papers in peer reviewed journals and international conferences.

Wei-Han Hung, a National Taiwan University (NTU) Masters graduate, is currently a Ph.D. student in the Department of Civil Engineering. His Masters focused on research in robot and visualization, which included the patrol path planning of a security robot, crane simulation, and virtual construction sites. Currently his research interest is developing path-planning methods for single and cooperative crane erections. He received "Best Student Presentation" award in CONVR, an international conference on construction applications of virtual reality.

Takeo Igarashi is an associate professor at the computer science department, at the University of Tokyo. He is also directing the Igarashi Design Interface Project, JST/ERATO. He received his Ph.D from the department of information engineering, the University of Tokyo in 2000. His research interest is in user interface in general and his current focus is on interaction techniques for 3D graphics and robots. He received the ACM SIGGRAPH 2006 significant new researcher award and the Katayanagi Prize in Computer Science.

Jinyung Jung is a Ph.D. candidate in the Department of Industrial Design at KAIST. He was trained as a product designer through the BS and MS programs in the same department and as an interaction design researcher through his involvement in various HRI design projects, such as Office robot design, Teaching assistant robot design, Childcare robot design, Intermediate interface design and Future robot computer design. His research interests lie in the fields of user experience design, user survey methodology. His Ph.D. research aims to establish a theoretical model to design motion of a humanoid robot for various speechless contexts.

David B. Kaber is a professor of Industrial and Systems Engineering at North Carolina State University and associate faculty in psychology. He received his PhD in industrial engineering from Texas Tech University in 1996. His current research interests include aircraft cockpit display design, computational modeling of pilot cognitive behavior, and driver situation awareness in hazardous conditions.

Shih-Chung Jessy Kang, a Stanford Ph.D. graduate, is currently an Assistant Professor in the Department of Civil Engineering at National Taiwan University. He used MSRDS in a research project from early 2006, half a year before the official release. From 2008, he started offering a hands-on robot development course for undergraduate students. In the same year, he obtained a teaching excellence award from NTU. From 2009, he started providing training courses for robot instructors in Taiwan to share his teaching experiences.

Myungsuk Kim holds a BFA from the Department of Applied Art, an MFA from the Department of Environmental Design, Hongik University, and a Ph. D. from the Department of Environmental Engineering, Osaka University, Japan. He is a professor in the Department of Industrial Design at KAIST and has been leading the Product and Environmental System Design Laboratory. His research interests are in the field of environmental system design, cultural design, emotional design, and robot design. He was the chairman of the Korean Society of Design Science (KSDS, 1997-2001) and Asian Design Congress (ADC, 1999-2001), which is a predecessor of the International Design Congress (IDC) and International Association of Society of Design Research (IASDR).

Sang-Hwan Kim is an assistant professor of Industrial and Manufacturing Systems Engineering at University of Michigan – Dearborn. He received his PhD in industrial and systems engineering from North Carolina State University in 2009. His current research interests include computational cognitive modeling, aircraft cockpit display design, vehicle human factors, and human computer interaction.

Yunkyung Kim is a Ph.D. candidate in the Department of Industrial Design at KAIST. She holds a BS and MS in the same department. Her research interests include mediating interface between human and computer, experience design, and

HRI. Her research method in HRI combines social psychology such as intergroup relations and psychological distance. Her research aim is to make designer design product with their intention and also to make user use product with various ways based on their cultural, environmental, and social attribution.

Kheng Lee Koay received his B.Sc. degree in robotics and automated systems and Ph.D. degree from the University of Plymouth, U.K. in 1997 and 2003, respectively. He is currently a Postdoctoral Researcher at the Adaptive Systems Research Group at the University of Hertfordshire, U.K. His research interests include Mobile Robotics, Human-Robot Interaction and Agent Migration. He was involved in the FP6 European project Cogniron and is currently working in the FP7 European project LIREC.

Bogdan Kreczmer received the M.Sc. degree in control in engineering from the Wrocław University of Technology, Poland, in 1986 and also M.Sc. degree in computer science, in 1988. He received the Ph.D. degree in robotics from the Wrocław University of Technology in 1997. He currently works as an assistant professor at the Institute of Computer Engineering, Control and Robotics, Department of Electronics, Wrocław University of Technology. His current research interests include the ultrasonic sensing and sensor data processing and mobile robot navigation.

Peter Liu received his B.Sc. and Ph.D. degrees in electrical engineering from Chung-Yuan Christian University, Chungli, Taiwan, in 1998 and 2002, respectively. In 2009, he received his MBA degree from National Taiwan University, Taipei, Taiwan. From 2002 to 2007, he was a senior researcher in the wireless BU of BenQ Corporations, Taiwan. He was an affiliated researcher from 2007 to 2009 in Robotics Lab, Department of Civil Engineering, National Taiwan University. From 2009, he joined the faculty of Department of Electrical Engineering, Tamkang University in Taipei, where he is currently an Assistant Professor.

Łukasz Małek received his M.Sc. degree in control engineering and robotics from the Wrocław University of Technology in 2005. Currently, he is a Ph.D. student at the same university. He also received an M.Sc. degree in mathematics from the Wrocław University of Technology in 2007. His research interests include motion planning as well as social robots design.

Kwangmyung Oh is a Ph.D. candidate in the Department of Industrial Design at KAIST. He was trained as a product designer through the BS and MS programs in the same department and as an interaction design researcher through his involvement in various HRI design projects, such as Ubiquitous Robot Companion (URC) design and emotional robot development. His research interests lie in the fields of user experience design, user-product relationship, and design prototyping for user research. His Ph.D. research focuses on the development of a theoretical experience model for robotic products to assist designers in better understanding the interaction dynamics between users and robotic products. In order to build this, he has been conducting HRI experiments through social attribute-centered approaches.

Michael A. Peshkin received the Ph.D. degree in physics from Carnegie-Mellon University in 1986. Since then, he has been a professor in the Department of Mechanical Engineering at Northwestern University where his research interests include physical human-robot interaction, haptics, rehabilitation robots, and sensors. He has served as associate editor for the IEEE Transactions on Robotics, chairs of various conference boards, and a member of the Robotics Council for the Robotics and Human Augmentation Program of the National Science Foundation. He is the co-inventor (with J. Edward Colgate) of the class of collaborative robots called "cobots", and holds various patents in robotics and sensors. Dr. Peshkin is also a cofounder of the companies Kinea Design, Cobotics, and Mako Surgical.

Iman Mohammad Rezazadeh received the B.Sc and M.Sc. degree in Biomedical Engineering from Science and Research Branch, Islamic Azad University (SRIAU), Tehran, Iran, in 2001 and 2003 respectively .He is now Ph.D. candidate and also affiliated to the clinical engineering group in the same institute since 2005. He is also researcher in Center of Advanced Researches and Technologies for Empowering and Mind Engineering (CARTEME) at SRIAU.

Iman's current research interests include the designing Human-Machine Interfaces for empowering and mind engineering, usability engineering, emotional intelligent and affective computing, cyber collaboration, virtual reality, and cybernetics.

Ehud Sharlin has been faculty at the University of Calgary's Computer Science Department since October 2004, following his position as faculty with the Human Interface Engineering Laboratory at Osaka University between 2003 and 2004. He is currently running the uTouch research group, and is a member of the University of Calgary Interactions Lab. His research interests are directed at interaction with physical objects and entities: human-robot interaction, tangible user interfaces, mixed reality and computer game interfaces.

Ehud completed his Ph.D. in Computing Science in July 2003 at the University of Alberta in Edmonton, Canada, under the supervision of Dr. Ben Watson and Dr. Jonathan Schaeffer, his M.Sc. in in 1997 (Magna Cum Laude), and his B.Sc. in 1990, both in Electrical and Computer Engineering from Ben-Gurion University, Israel, under the supervision of Dr. Jonathan Molcho..

Between 1991 and 1998 Ehud worked as a senior researcher and research director with several Israeli R&D labs, his main research themes during this period were image processing, computer vision and tracking algorithms, electro-optical design, numerical simulation, and user studies.

Dag Sverre Syrdal received his BSc in Psychology from Queen's University Belfast in 2001, and his MSc in Research Methods and Data Analysis in Psychology from the University of Hertfordshire in 2002. He joined the Adaptive Systems Research Group at the University of Hertfordshire in 2006, where his work has centred around the planning, execution and analysis of user studies within the Cogniron and LIREC projects. His research interests include Human-Robot Interaction, with an emphasis on socially acceptable interactions in human-centred environments as well as the role of Individual Differences in HRI.

Xiangyu Wang is a Senior Lecturer in the Faculty of Built Environment, at The University of New South Wales, Australia. He obtained his PhD degree in Civil Engineering at Purdue University in 2005. Dr Wang's work features highly Interdisciplinary Research across Design, Computer Engineering, Construction, and Human Factors. His specific research interests include Virtual Environments for Design, Human-Computer Interactions, Computer-Supported Cooperative Work, and Construction Automation and Robotics. He has published over 70 refereed articles in a wide range of highly recognised international journals and conferences (ASCE, IEEE, ACM, etc.). He was awarded a US National Science Foundation grant to investigate Skill Development through Virtual Technologies.

Xuexhong Wang is Ph.D. candidate and graduate research assistant in Industrial and Systems Engineering at North Carolina State University. He received bachelors degrees in civil engineering and industrial engineering from South China University of Technology in 2004. His current research interests include gaze-based human-computer interaction, pattern recognition and optimum control.

James (Jim) E. Young is currently a Ph.D. Candidate at the University of Calgary, under the supervision of Dr. Ehud Sharlin (University of Calgary) and Dr. Takeo Igarashi (Tokyo University). James is a researcher with both the University of Calgary's uTouch research group (part of the Interactions Lab, Calgary, Canada) and JST Erato Laboratory (Tokyo, Japan).

His research focuses on human-robot interaction, tangible interfaces, human-computer interaction, and computer vision. James has led several unique research projects exploring the use of mixed-reality techniques in human-robots interaction, and designing robotic behaviors based on programming-by-demonstration.

James completed his B.Sc. at Vancouver Island University in 2005. His senior project, titled "Space-Model Machine" and directed by Dr. David Wessels, was on modeling logic problems in multi-agent systems. James is currently pursuing his Ph.D. in Computer Science at the University of Calgary and is expected to graduate in early 2010.

Jie Zhu is currently a Ph.D candidate at the Faculty of Architecture, Design and Planning, the University of Sydney, Australia, under the supervision of Dr. Xiangyu Wang. Jie Zhu has two master degrees in IT. His main research interests include robotics study in mixed reality environments, situation awareness, collaborative virtual environments, etc.

Author Index

Index

K

L

M

N

O

P

Q

R

S

T

U

V

W