Julie A. Jacko (Ed.)

Human-Computer Interaction

Towards Mobile and Intelligent Interaction Environments

LNCS 6763

14th International Conference, HCI International 2011
Orlando, FL, USA, July 2011
Proceedings, Part III

3 Part III

HCI 2011 INTERNATIONAL

Springer

Lecture Notes in Computer Science 6763

Julie A. Jacko (Ed.)

Human-Computer Interaction

Towards Mobile and Intelligent Interaction Environments

14th International Conference, HCI International 2011
Orlando, FL, USA, July 9-14, 2011
Proceedings, Part III

Springer

Volume Editor

Julie A. Jacko
University of Minnesota
School of Public Health and Institute for Health Informatics
1260 Mayo (MMC 807), 420 Delaware Street S.E.
Minneapolis, MN 55455, USA
E-mail: jacko@umn.edu

ISSN 0302-9743 e-ISSN 1611-3349
ISBN 978-3-642-21615-2 e-ISBN 978-3-642-21616-9
DOI 10.1007/978-3-642-21616-9
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2011929076

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

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

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

Printed on acid-free paper

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

Foreword

The 14th International Conference on Human–Computer Interaction, HCI International 2011, was held in Orlando, Florida, USA, July 9–14, 2011, jointly with the Symposium on Human Interface (Japan) 2011, the 9th International Conference on Engineering Psychology and Cognitive Ergonomics, the 6th International Conference on Universal Access in Human–Computer Interaction, the 4th International Conference on Virtual and Mixed Reality, the 4th International Conference on Internationalization, Design and Global Development, the 4th International Conference on Online Communities and Social Computing, the 6th International Conference on Augmented Cognition, the Third International Conference on Digital Human Modeling, the Second International Conference on Human-Centered Design, and the First International Conference on Design, User Experience, and Usability.

A total of 4,039 individuals from academia, research institutes, industry and governmental agencies from 67 countries submitted contributions, and 1,318 papers that were judged to be of high scientific quality were included in the program. These papers address the latest research and development efforts and highlight the human aspects of design and use of computing systems. The papers accepted for presentation thoroughly cover the entire field of human–computer interaction, addressing major advances in knowledge and effective use of computers in a variety of application areas.

This volume, edited by Julie A. Jacko, contains papers in the thematic area of human–computer interaction (HCI), addressing the following major topics:

- Mobile interaction
- Interaction in intelligent environments
- Orientation and navigation
- In-vehicle interaction
- Social and environmental issues in HCI
- Emotions in HCI

The remaining volumes of the HCI International 2011 Proceedings are:

- Volume 1, LNCS 6761, Human–Computer Interaction—Design and Development Approaches (Part I), edited by Julie A. Jacko
- Volume 2, LNCS 6762, Human–Computer Interaction—Interaction Techniques and Environments (Part II), edited by Julie A. Jacko
- Volume 4, LNCS 6764, Human–Computer Interaction—Users and Applications (Part IV), edited by Julie A. Jacko
- Volume 5, LNCS 6765, Universal Access in Human–Computer Interaction—Design for All and eInclusion (Part I), edited by Constantine Stephanidis
- Volume 6, LNCS 6766, Universal Access in Human–Computer Interaction—Users Diversity (Part II), edited by Constantine Stephanidis

- Volume 7, LNCS 6767, Universal Access in Human–Computer Interaction—Context Diversity (Part III), edited by Constantine Stephanidis
- Volume 8, LNCS 6768, Universal Access in Human–Computer Interaction—Applications and Services (Part IV), edited by Constantine Stephanidis
- Volume 9, LNCS 6769, Design, User Experience, and Usability—Theory, Methods, Tools and Practice (Part I), edited by Aaron Marcus
- Volume 10, LNCS 6770, Design, User Experience, and Usability—Understanding the User Experience (Part II), edited by Aaron Marcus
- Volume 11, LNCS 6771, Human Interface and the Management of Information—Design and Interaction (Part I), edited by Michael J. Smith and Gavriel Salvendy
- Volume 12, LNCS 6772, Human Interface and the Management of Information—Interacting with Information (Part II), edited by Gavriel Salvendy and Michael J. Smith
- Volume 13, LNCS 6773, Virtual and Mixed Reality—New Trends (Part I), edited by Randall Shumaker
- Volume 14, LNCS 6774, Virtual and Mixed Reality—Systems and Applications (Part II), edited by Randall Shumaker
- Volume 15, LNCS 6775, Internationalization, Design and Global Development, edited by P.L. Patrick Rau
- Volume 16, LNCS 6776, Human-Centered Design, edited by Masaaki Kurosu
- Volume 17, LNCS 6777, Digital Human Modeling, edited by Vincent G. Duffy
- Volume 18, LNCS 6778, Online Communities and Social Computing, edited by A. Ant Ozok and Panayiotis Zaphiris
- Volume 19, LNCS 6779, Ergonomics and Health Aspects of Work with Computers, edited by Michelle M. Robertson
- Volume 20, LNAI 6780, Foundations of Augmented Cognition: Directing the Future of Adaptive Systems, edited by Dylan D. Schmorrow and Cali M. Fidopiastis
- Volume 21, LNAI 6781, Engineering Psychology and Cognitive Ergonomics, edited by Don Harris
- Volume 22, CCIS 173, HCI International 2011 Posters Proceedings (Part I), edited by Constantine Stephanidis
- Volume 23, CCIS 174, HCI International 2011 Posters Proceedings (Part II), edited by Constantine Stephanidis

I would like to thank the Program Chairs and the members of the Program Boards of all Thematic Areas, listed herein, for their contribution to the highest scientific quality and the overall success of the HCI International 2011 Conference.

In addition to the members of the Program Boards, I also wish to thank the following volunteer external reviewers: Roman Vilimek from Germany, Ramalingam Ponnusamy from India, Si Jung "Jun" Kim from the USA, and Ilia Adami, Iosif Klironomos, Vassilis Kouroumalis, George Margetis, and Stavroula Ntoa from Greece.

This conference would not have been possible without the continuous support and advice of the Conference Scientific Advisor, Gavriel Salvendy, as well as the dedicated work and outstanding efforts of the Communications and Exhibition Chair and Editor of HCI International News, Abbas Moallem.

I would also like to thank for their contribution toward the organization of the HCI International 2011 Conference the members of the Human–Computer Interaction Laboratory of ICS-FORTH, and in particular Margherita Antona, George Paparoulis, Maria Pitsoulaki, Stavroula Ntoa, Maria Bouhli and George Kapnas.

July 2011 Constantine Stephanidis

Organization

Ergonomics and Health Aspects of Work with Computers

Program Chair: Michelle M. Robertson

Arne Aarås, Norway
Pascale Carayon, USA
Jason Devereux, UK
Wolfgang Friesdorf, Germany
Martin Helander, Singapore
Ed Israelski, USA
Ben-Tzion Karsh, USA
Waldemar Karwowski, USA
Peter Kern, Germany
Danuta Koradecka, Poland
Nancy Larson, USA
Kari Lindström, Finland
Brenda Lobb, New Zealand
Holger Luczak, Germany
William S. Marras, USA
Aura C. Matias, Philippines
Matthias Rötting, Germany
Michelle L. Rogers, USA
Dominique L. Scapin, France
Lawrence M. Schleifer, USA
Michael J. Smith, USA
Naomi Swanson, USA
Peter Vink, The Netherlands
John Wilson, UK

Human Interface and the Management of Information

Program Chair: Michael J. Smith

Hans-Jörg Bullinger, Germany
Alan Chan, Hong Kong
Shin'ichi Fukuzumi, Japan
Jon R. Gunderson, USA
Michitaka Hirose, Japan
Jhilmil Jain, USA
Yasufumi Kume, Japan
Mark Lehto, USA
Hirohiko Mori, Japan
Fiona Fui-Hoon Nah, USA
Shogo Nishida, Japan
Robert Proctor, USA
Youngho Rhee, Korea
Anxo Cereijo Roibás, UK
Katsunori Shimohara, Japan
Dieter Spath, Germany
Tsutomu Tabe, Japan
Alvaro D. Taveira, USA
Kim-Phuong L. Vu, USA
Tomio Watanabe, Japan
Sakae Yamamoto, Japan
Hidekazu Yoshikawa, Japan
Li Zheng, P. R. China

Human–Computer Interaction

Program Chair: Julie A. Jacko

Sebastiano Bagnara, Italy
Sherry Y. Chen, UK
Marvin J. Dainoff, USA
Jianming Dong, USA
John Eklund, Australia
Xiaowen Fang, USA
Ayse Gurses, USA
Vicki L. Hanson, UK
Sheue-Ling Hwang, Taiwan
Wonil Hwang, Korea
Yong Gu Ji, Korea
Steven A. Landry, USA
Gitte Lindgaard, Canada
Chen Ling, USA
Yan Liu, USA
Chang S. Nam, USA
Celestine A. Ntuen, USA
Philippe Palanque, France
P.L. Patrick Rau, P.R. China
Ling Rothrock, USA
Guangfeng Song, USA
Steffen Staab, Germany
Wan Chul Yoon, Korea
Wenli Zhu, P.R. China

Engineering Psychology and Cognitive Ergonomics

Program Chair: Don Harris

Guy A. Boy, USA
Pietro Carlo Cacciabue, Italy
John Huddlestone, UK
Kenji Itoh, Japan
Hung-Sying Jing, Taiwan
Wen-Chin Li, Taiwan
James T. Luxhøj, USA
Nicolas Marmaras, Greece
Sundaram Narayanan, USA
Mark A. Neerincx, The Netherlands
Jan M. Noyes, UK
Kjell Ohlsson, Sweden
Axel Schulte, Germany
Sarah C. Sharples, UK
Neville A. Stanton, UK
Xianghong Sun, P.R. China
Andrew Thatcher, South Africa
Matthew J.W. Thomas, Australia
Mark Young, UK
Rolf Zon, The Netherlands

Universal Access in Human–Computer Interaction

Program Chair: Constantine Stephanidis

Julio Abascal, Spain
Ray Adams, UK
Elisabeth André, Germany
Margherita Antona, Greece
Chieko Asakawa, Japan
Christian Bühler, Germany
Jerzy Charytonowicz, Poland
Pier Luigi Emiliani, Italy
Michael Fairhurst, UK
Dimitris Grammenos, Greece
Andreas Holzinger, Austria
Simeon Keates, Denmark
Georgios Kouroupetroglou, Greece
Sri Kurniawan, USA
Patrick M. Langdon, UK
Seongil Lee, Korea

Zhengjie Liu, P.R. China
Klaus Miesenberger, Austria
Helen Petrie, UK
Michael Pieper, Germany
Anthony Savidis, Greece
Andrew Sears, USA
Christian Stary, Austria
Hirotada Ueda, Japan
Jean Vanderdonckt, Belgium
Gregg C. Vanderheiden, USA
Gerhard Weber, Germany
Harald Weber, Germany
Panayiotis Zaphiris, Cyprus

Virtual and Mixed Reality

Program Chair: Randall Shumaker

Pat Banerjee, USA
Mark Billinghurst, New Zealand
Charles E. Hughes, USA
Simon Julier, UK
David Kaber, USA
Hirokazu Kato, Japan
Robert S. Kennedy, USA
Young J. Kim, Korea
Ben Lawson, USA
Gordon McK Mair, UK
David Pratt, UK
Albert "Skip" Rizzo, USA
Lawrence Rosenblum, USA
Jose San Martin, Spain
Dieter Schmalstieg, Austria
Dylan Schmorrow, USA
Kay Stanney, USA
Janet Weisenford, USA
Mark Wiederhold, USA

Internationalization, Design and Global Development

Program Chair: P.L. Patrick Rau

Michael L. Best, USA
Alan Chan, Hong Kong
Lin-Lin Chen, Taiwan
Andy M. Dearden, UK
Susan M. Dray, USA
Henry Been-Lirn Duh, Singapore
Vanessa Evers, The Netherlands
Paul Fu, USA
Emilie Gould, USA
Sung H. Han, Korea
Veikko Ikonen, Finland
Toshikazu Kato, Japan
Esin Kiris, USA
Apala Lahiri Chavan, India
James R. Lewis, USA
James J.W. Lin, USA
Rungtai Lin, Taiwan
Zhengjie Liu, P.R. China
Aaron Marcus, USA
Allen E. Milewski, USA
Katsuhiko Ogawa, Japan
Oguzhan Ozcan, Turkey
Girish Prabhu, India
Kerstin Röse, Germany
Supriya Singh, Australia
Alvin W. Yeo, Malaysia
Hsiu-Ping Yueh, Taiwan

Online Communities and Social Computing

Program Chairs: A. Ant Ozok, Panayiotis Zaphiris

Chadia N. Abras, USA
Chee Siang Ang, UK
Peter Day, UK
Fiorella De Cindio, Italy
Heidi Feng, USA
Anita Komlodi, USA
Piet A.M. Kommers, The Netherlands
Andrew Laghos, Cyprus
Stefanie Lindstaedt, Austria
Gabriele Meiselwitz, USA
Hideyuki Nakanishi, Japan
Anthony F. Norcio, USA
Ulrike Pfeil, UK
Elaine M. Raybourn, USA
Douglas Schuler, USA
Gilson Schwartz, Brazil
Laura Slaughter, Norway
Sergei Stafeev, Russia
Asimina Vasalou, UK
June Wei, USA
Haibin Zhu, Canada

Augmented Cognition

Program Chairs: Dylan D. Schmorrow, Cali M. Fidopiastis

Monique Beaudoin, USA
Chris Berka, USA
Joseph Cohn, USA
Martha E. Crosby, USA
Julie Drexler, USA
Ivy Estabrooke, USA
Chris Forsythe, USA
Wai Tat Fu, USA
Marc Grootjen, The Netherlands
Jefferson Grubb, USA
Santosh Mathan, USA
Rob Matthews, Australia
Dennis McBride, USA
Eric Muth, USA
Mark A. Neerincx, The Netherlands
Denise Nicholson, USA
Banu Onaral, USA
Kay Stanney, USA
Roy Stripling, USA
Rob Taylor, UK
Karl van Orden, USA

Digital Human Modeling

Program Chair: Vincent G. Duffy

Karim Abdel-Malek, USA
Giuseppe Andreoni, Italy
Thomas J. Armstrong, USA
Norman I. Badler, USA
Fethi Calisir, Turkey
Daniel Carruth, USA
Keith Case, UK
Julie Charland, Canada
Yaobin Chen, USA
Kathryn Cormican, Ireland
Daniel A. DeLaurentis, USA
Yingzi Du, USA
Okan Ersoy, USA
Enda Fallon, Ireland
Yan Fu, P.R. China
Afzal Godil, USA

Ravindra Goonetilleke, Hong Kong
Anand Gramopadhye, USA
Lars Hanson, Sweden
Pheng Ann Heng, Hong Kong
Bo Hoege, Germany
Hongwei Hsiao, USA
Tianzi Jiang, P.R. China
Nan Kong, USA
Steven A. Landry, USA
Kang Li, USA
Zhizhong Li, P.R. China
Tim Marler, USA
Ahmet F. Ozok, Turkey
Srinivas Peeta, USA
Sudhakar Rajulu, USA
Matthias Rötting, Germany
Matthew Reed, USA
Johan Stahre, Sweden
Mao-Jiun Wang, Taiwan
Xuguang Wang, France
Jingzhou (James) Yang, USA
Gulcin Yucel, Turkey
Tingshao Zhu, P.R. China

Human-Centered Design

Program Chair: Masaaki Kurosu

Julio Abascal, Spain
Simone Barbosa, Brazil
Tomas Berns, Sweden
Nigel Bevan, UK
Torkil Clemmensen, Denmark
Susan M. Dray, USA
Vanessa Evers, The Netherlands
Xiaolan Fu, P.R. China
Yasuhiro Horibe, Japan
Jason Huang, P.R. China
Minna Isomursu, Finland
Timo Jokela, Finland
Mitsuhiko Karashima, Japan
Tadashi Kobayashi, Japan
Seongil Lee, Korea
Kee Yong Lim, Singapore
Zhengjie Liu, P.R. China
Loïc Martínez-Normand, Spain
Monique Noirhomme-Fraiture, Belgium
Philippe Palanque, France
Annelise Mark Pejtersen, Denmark
Kerstin Röse, Germany
Dominique L. Scapin, France
Haruhiko Urokohara, Japan
Gerrit C. van der Veer, The Netherlands
Janet Wesson, South Africa
Toshiki Yamaoka, Japan
Kazuhiko Yamazaki, Japan
Silvia Zimmermann, Switzerland

Design, User Experience, and Usability

Program Chair: Aaron Marcus

Ronald Baecker, Canada
Barbara Ballard, USA
Konrad Baumann, Austria
Arne Berger, Germany
Randolph Bias, USA
Jamie Blustein, Canada
Ana Boa-Ventura, USA
Lorenzo Cantoni, Switzerland
Sameer Chavan, Korea
Wei Ding, USA
Maximilian Eibl, Germany
Zelda Harrison, USA

Rüdiger Heimgärtner, Germany
Brigitte Herrmann, Germany
Sabine Kabel-Eckes, USA
Kaleem Khan, Canada
Jonathan Kies, USA
Jon Kolko, USA
Helga Letowt-Vorbek, South Africa
James Lin, USA
Frazer McKimm, Ireland
Michael Renner, Switzerland
Christine Ronnewinkel, Germany
Elizabeth Rosenzweig, USA
Paul Sherman, USA
Ben Shneiderman, USA
Christian Sturm, Germany
Brian Sullivan, USA
Jaakko Villa, Finland
Michele Visciola, Italy
Susan Weinschenk, USA

HCI International 2013

The 15th International Conference on Human–Computer Interaction, HCI International 2013, will be held jointly with the affiliated conferences in the summer of 2013. It will cover a broad spectrum of themes related to human–computer interaction (HCI), including theoretical issues, methods, tools, processes and case studies in HCI design, as well as novel interaction techniques, interfaces and applications. The proceedings will be published by Springer. More information about the topics, as well as the venue and dates of the conference, will be announced through the HCI International Conference series website: http://www.hci-international.org/

General Chair
Professor Constantine Stephanidis
University of Crete and ICS-FORTH
Heraklion, Crete, Greece
Email: cs@ics.forth.gr

Table of Contents – Part III

Part I: Mobile Interaction

Part II: Interaction in Intelligent Environments

Part III: Orientation and Navigation

Part IV: In-Vehicle Interaction

Part V: Social and Environmental Issues in HCI

Part VI: Emotions in HCI

Part I
Mobile Interaction

Field to File: A Tool for Activity Documentation Work in Remote Mobility Environments

Raúl Casillas[1] and Alberto L. Morán[2]

[1] Facultad de Ingeniería y Negocios San Quintín, UABC, Ensenada, B.C., México
[2] Facultad de Ciencias, UABC, Ensenada, B.C., México
{rcasillas,alberto.moran}@uabc.edu.mx

Abstract. Previous studies have identified that activity documentation of mobile workers is carried out during three different documentation stages: before (pre-), during (at-) and after (post-) fieldwork activities. The first stage includes fieldwork planning and equipment preparation for the job, the second includes both an informative meeting about the activities that will take place during fieldwork and the fieldwork activities themselves, while the third and final stage includes the preparation and delivery of a report generated with the information collected during fieldwork. Based on this understanding, these studies have proposed a number of features to include in a tool that seeks to support the documentation of this type of work, including support for fixed and mobile modes of work and seamless access to information across stages and modes of work. In this paper we present the design and evaluation of a prototype tool named Field to File, which seeks to support activity documentation of workers in remote mobile environments that have not been augmented with technology.

Keywords: Activity documentation work, remote mobility environments, mobile computing, Field to File.

1 Introduction

Activity documentation is a critical part of the work of many professionals. Documents are used as a means to store personal information, remind things to do, convey and generate new meaning, and mediate contact among people.

Hertzum [1] identifies six roles documents play in professional work, including their use as i) personal files, ii) as reminders of things to do, iii) to share information with someone, yet withholding it from others, iv) to convey meaning, v) to generate new meaning, and vi) to mediate contacts among people.

In a previous study [2], we characterized application domains from the perspective of "worker's mobility" and "scale of the environment" where the documentation activity occurs. The proposed classification includes: i) *Workers with micro mobility in "at-hand" bounded environments.* It refers to the mobility of individuals performing activities, carrying and manipulating artifacts, in circumscribed places; ii) *Workers with local mobility within an environment with a number of workspaces.* It takes place

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 3–12, 2011.

in settings where actors are constantly on the move to get ahead with their work; and iii) *Workers with remote mobility in wide open environments*. It refers to workers that travel to particular sites (field), and once there, move throughout the place (even walking considerable distances) in order to perform their work.

In this paper, we are interested in the last kind of workers, whose work is characterized by periods of micro- and local- mobility work at a local setting, and periods of remote mobility work in the field. Particularly, we are interested on providing support for activity documentation of workers in remote mobility environments (RME).

The rest of the paper is organized as follows: section 2 introduces activity documentation work in RME as well as a set of design insights for the construction of tools for activity documentation work in this kind of environments. Section 3 takes up on the design and implementation of Field to File, a tool to provide explicit support for activity documentation work in RME. Section 4 presents the results of an initial evaluation of the proposed tool, while Section 5 presents and discusses related work. Finally, Section 6 presents our concluding remarks along with directions of future work.

2 Activity Documentation Work in RME

In [2] we identified the main activities of remote mobility workers and arranged these activities following a sequential life-line that is more or less followed (and repeated) each time fieldwork activity in RME is conducted (see Fig. 1). These activities were also arranged to denote the activities that are performed before (pre-), during (at-) and after (post-) fieldwork activities.

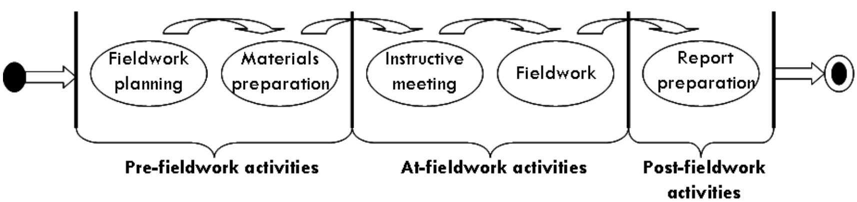


Fig. 1. Activities performed by workers before, during and after fieldwork activities

Also, relevant challenges which workers may face while performing activity documentation work in RME were identified, including [2]:

- Imbalance in the effort spent in the phases of activity documentation work.
- Need to access information and documentation products of previous phases.
- Information capture using non-digital media.
- Fragmented nature of the information captured in different media and lack of effective mechanisms for linking related information.
- Need to share the collected information among colleagues, and to collaborate in the creation of deliverables.

Considering these situations faced by workers, we also identified a set of design implications (see Table 1) that a technological tool that seeks to support activity documentation work in RME should address [2].

Table 1. Identified design insights for the provision of support for activity documentation work in RME

Design Insights
1.- Support for fixed and mobile modes of work, according to the needs of activity documentation work stages.
2.- Seamless access to information across the phases and modes of activity documentation work.
3.- Digital capture of activity documentation information, so that even transient information (e.g. conversations between colleagues) could be persistently documented.
4.- Automatic linking of different pieces of related information, even at capture time, to avoid additional classification work in the Post-activity documentation stage.
5.- Information sharing among colleagues, and collaborative creation of individual and group deliverables, regardless of the time and place where participants are.

In the following sections, we describe the design, developlment and initial evaluation of a prototype tool that by addressing most of the proposed design insights seeks to provide support for activity documentation work in RME.

3 Field to File: A Tool for Activity Documentation Work in RME

Field to File is a tool that provides support for activity documentation work in RME. To do so, it considers essential features not only to support a single activity, but also to create and integrate fixed and mobile services for information capture, classification, retrieval and processing, which are available and interoperable throughout the complete life-cycle of activity documentation work in RME. The current implementation of Field to File uses off-the-shelf components, some of them extended to support specific features of our design.

3.1 Mobile Field to File

Mobile Field to File runs on a TabletPC Mobile Computing F5 with Windows XP TabletPC Edition and MS-Office 2003 installed; and with integrated digital camera, microphone and Wi-Fi support. The GUI interface and main components of Mobile Field to File are shown in Fig. 2. A brief description of these components follows.

Note-taking component. It is a MS Word component that has been embedded into the Mobile Field to File application. It provides the complete functionality of the word processor application to be accessible inside the integrated documentation environment. To simplify the documentation process, functionality to automatically name the note and define its storage location was added. This way, users only have to type the note and save it on exiting the note-taking feature.

Picture taking component. It utilizes the Windows Image Acquisition (IWA) library to take advantage of the device's integrated digital camera. An interface to take multiple pictures and automatically name and set the storage location of files was programmed. This way, users only have to take the picture and the system automatically names it and saves it on exiting the component.

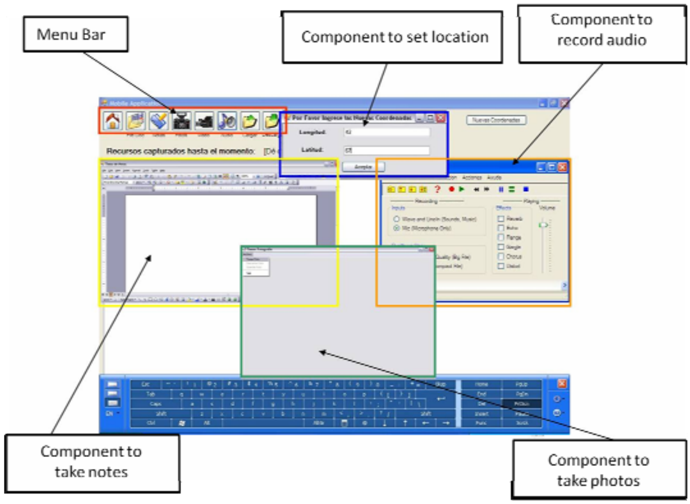


Fig. 2. Main interface and system components of Field to File - Mobile

Audio Recording component. This is the open source Sound Recorder Plus component that has been embedded into the integrated documentation environment to provide audio recording capabilities using the device's integrated microphone. It has been extended with functionality to automatically name the audio file and to define its storage location. This way, users only have to start and stop recording a conversation, and the system automatically names it and saves it on exiting the component.

Setting location component. In this implementation, the component allows users to define the current location manually, by setting the UTM coordinate values from an external GPS system. These values are used to generate unique names for information elements generated or captured at specific locations.

File Manager component. This module is in charge of automatically generating the names for the captured or generated information elements, and storing them in a specific file system path. It is responsible for assuring name uniqueness and of automatically transferring information from the mobile to the desktop subsystem.

3.2 Desktop Field to File

The GUI interface and main components of Desktop Field to File are shown in Fig. 3. A brief description of these components follows.

Report Creation component. It is a MS Word component that was embedded into the Desktop Field to File tool. It provides the complete functionality of the word processor application inside the integrated documentation environment. To simplify the report generation, functionality to automatically insert specific information elements directly from the repositories was added.

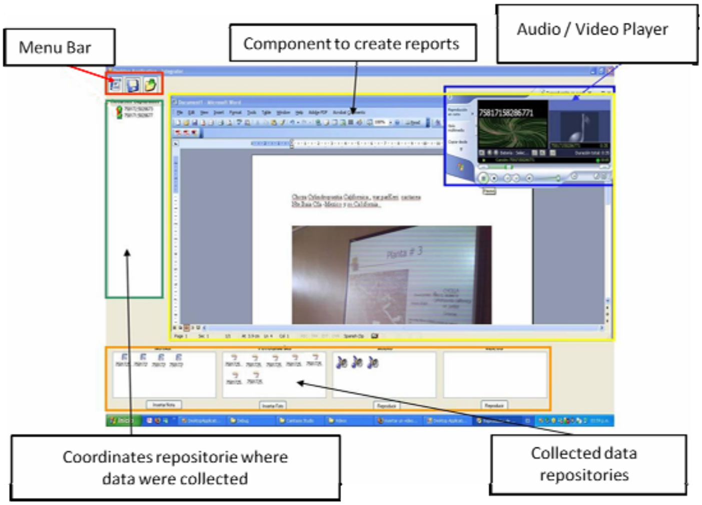


Fig. 3. Main interface and system components of Desktop Field to File

Information Aggregates repository component. It provides an entry point to the collection of information elements under the form of information aggregates. In it, entries represent the aggregates by location. This way, users select particular locations and the system loads all the captured information elements associated to that location, which are then presented in the notes, pictures and audio sub-repository views.

Text Notes repository component. It is automatically populated with all generated text notes associated with the location specified at the information aggregates level. Here, text notes can be selected, previewed using specified viewers, and automatically inserted into the current cursor location in the report creation component.

Pictures repository component. Similarly, it is automatically filled with captured pictures associated with the location specified at the information aggregates level. Here, pictures can be selected, previewed using specified viewers, and automatically inserted into the current cursor location in the report creation component.

Audio files repository component. In a similar way, this component is automatically filled with all audio files captured in the location specified at the information aggregate level. However, in this case, audio files can only be previewed using the specified player. Users may type information in the report creation component while playing the audio file in the provided player.

File Manager component. This component receives the information elements that are transferred from the Mobile Field to File subsystem and stores them in the desktop file repository. It manages the file repository in order to automatically store information aggregates by location, and to automatically populate the text notes, pictures and audio files repository components. This way, the user only has to deal with collection site's locations, instead of with the names and types of files of the captured or generated information resources.

4 Initial Evaluation of Field to File

We conducted a controlled evaluation experiment to assess whether the provided support for activity documentation work of our proposed tool was adequate and fit the needs of our prospective users.

Objective. The aim of the evaluation was to determine if the proposed tool would allow workers to improve their performance to conduct activity documentation work.

Hypotheses. The following hypotheses guided the evaluation process:

H0: No difference exists in terms of performance to conduct activity documentation work in RME between using the traditional tools (paper and pencil, digital camera, etc.) and the Field to File tool;

H1: The time required to capture information about the objects of study will be less when the task is conducted in the traditional way (i.e. using paper and pencil, digital camera, etc.) than when it is conducted using the Field to File tool;

H2: The time required to transfer and transcript the captured data and to write the deliverable report will be less when performing it using the Field to File tool than when it is conducted in the traditional way (using paper, pencil, digital camera, etc.).

Methodology. We conducted a controlled experiment where a group of biologists executed a set of predetermined documentation work activities from the pre-, at- and post-fieldwork stages. We used a "within-subjects" paradigm (all users participated in all modalities) to compare their performance to document their work in their usual way (not using the tool) vs. using the proposed tool. To avoid learning, five participants executed their activities firstly without the tool and later using the tool, while the other five executed their activities in a reverse order (firstly using the tool and then without using it).

Participants. Participants in the experiment were 3 students from the Biology undergraduate program and 7 students from the Management of Ecosystems in Arid Zones master's program, both from the Faculty of Sciences, Autonomous University of Baja California, at Ensenada, B.C., México. There were 3 male and 7 female participants. Their ages ranged from 22 to 35 years old. All of them had at least 3 years of experience conducting activity documentation work in RME and at least 6 years of experience using computer equipment and documentation applications.

Setting. The experiment was conducted at a multi-purpose lab prepared with a pre- and post-fieldwork activity setting and an at-fieldwork activity setting (see Fig. 4). The pre- and post-fieldwork activity setting consisted of a table, chair, printer and desktop computer with both traditional office productivity tools (MS Office, Picture browser, media player, etc.) and the Desktop Field to File application. Wi-Fi connectivity was also provided. The at-fieldwork activity setting consisted of a smart board where pictures and descriptions of the objects to be documented were presented. Also, subjects had access to paper notebooks, pen, pencil, digital camera, GPS device, and digital sound recorder which are tools they usually bring with them to conduct activity documentation work in RME. They also had access to the TabletPC Mobile Computing F5 with the Mobile Field to File application installed.

Fig. 4. Setting the meeting room where the experiment was conducted

***Procedure*.** Initially, subjects were introduced to the experiment, and asked to sign a participation consent form. Then they were asked to answer an on-entry questionnaire aimed at obtaining demographic data, and determining their experience on activity documentation work in RME and on the use of computer equipment, documentation applications and other digital equipment. Later, subjects were provided with printed instructions for the execution of three activities: i) gather information regarding the types of plants they should identify and study during the at-fieldwork stage – they used the desktop computer to find, print or transfer documentation related to these plants, ii) document their work studying these two plants during fieldwork – they were presented with pictures of the plants and with videos of experts describing the features of the plants and used either the traditional support or the provided Mobile Field to File tool as required for the modality, and, iii) they were asked to create a report with the captured information either using their traditional tools or the provided desktop Field to File tool. They were required to include at least a picture of each plant, and to provide specific data about them. Before performing these tasks in the "using Field to File tool" modality, participants received a 10 minute instructive session on the use of the tool. Participants were not instructed on how to perform tasks in the "traditional" modality.

***Results*.** As stated earlier, we observed and measured how participants conducted activity documentation work in two different modalities during a controlled experiment, and analyzed each of the previously defined working hypotheses against data from the experiment. The results for each of the scenarios follow.

4.1 Time Spent on Data Capture

Fig. 5 shows the time (in seconds) it took each participant to complete data capture, both in the "traditional way" and with the use of the proposed system. The average time to complete the activity was higher when using Field to File, however the standard deviation was greater when it was not used, so we performed a *t-test* with 99 degrees of confidence to test whether the difference is significant (see Table 2).

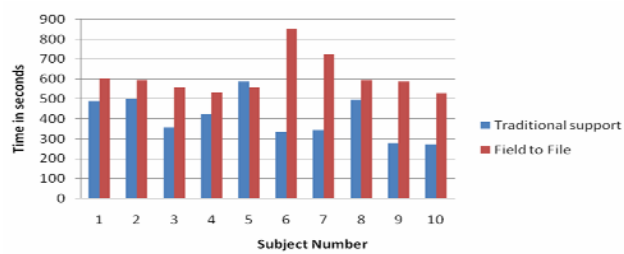


Fig. 5. Time required to complete data capture

Table 2. Analysis of the time required for data capture

	Traditional support	*Field to File*
Mean	408	612.4
Variance	11432.88889	9949.6
t Stat	-4.420300189	
t Critical one-tail	2.552379618	

As it can be seen in Table 2, the value of the *t* statistic (-4.420300189) is less than the critical value of *t* for one tail (2.552379618), therefore we can say that there is no significant difference to support *H1*, i.e. it is not supported that data capture using the traditional tools takes less time than using the Field to File tool (*H1* is rejected).

4.2 Time Spent on Data Transcription and Report Generation

Fig. 6 presents the time (in seconds) required by each participant to complete the transcription and transfer of data and its integration into a report, both in the "traditional way" and with the use of Field to File.

As can be seen, the average time and standard deviation required to complete the activity is higher when using traditional tools than when using Field to File. In order to establish whether the difference is significant, we carried out a *t-test* with 99 degrees of confidence (see Table 3).

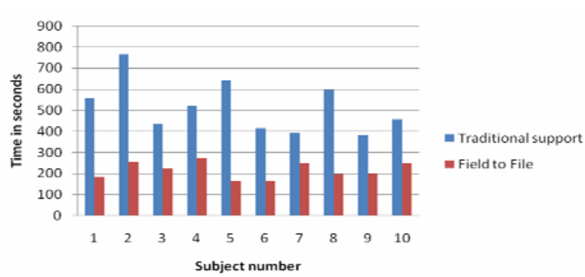


Fig. 6. Time required to complete data transcription and report creation

As it can be seen in Table 3, the value of the *t* statistic (7.263302927) is greater than the critical value of *t* for one tail (2.718079183), which leads to establish that there is a significant difference to support that to transfer and transcript the collected data and generating a report with Field to File requires less time than when using traditional tools (H2 is accepted).

Table 3. Analysis of time required for data transcription and report generation

	Traditional support	*Field to File*
Mean	516.6	216.2
Variance	15544.26667	1561.066667
t Stat	7.263302927	
t Critical one-tail	2.718079183	

5 Related Work

Support for Activity Documentation Work has been widely studied in diverse environments by means of capture and access support systems. Systems closer in spirit to Field to File include the Ambient Wood [3], the SenseCam [4], and the ButterflyNet [5] projects. The Ambient Wood Project aims to facilitate indoor and outdoor learning experiences of students using custom infrastructure in both kinds of environments. However, in contrast to Field to File, it requires that technology (e.g. sensors and speakers hidden in the environment) should be already installed both indoors and outdoors in order for it to work properly.

The SenseCam project [4], proposes the use of the SenseCam (a context-aware digital camera) to visually document people's activities to later rebuild their day and help them remember what they did. In contrast with our system, the SenseCam project does not focuses on the use of text notes and audio files as resources to document people's activities, nor provides support for the creation of a written deliverable.

Finally, the ButterflyNet project [5] proposes to provide support for the activity documentation work of a group of biologists in RME. They do so by augmenting biologists' pen and paper practices by introducing the use of a digital pen, and thus not altering how they perform their documenting activity. A system to integrate the gathered data and to generate a written report is also provided. This way, the digital pen and the provided system allow for the inclusion of written commands at capture time, which are later interpreted and that allow for the inclusion of the text notes and pictures generated and captured during the at-fieldwork stage. However, in contrast to Field to File, if ButterflyNet users would like to include and edit their text notes in their written report during the post-fieldwork stage, they must transcribe them, as what they actually obtain and insert are just digital images of the notes. In addition, ButteflyNet lacks the support to gather, store and bring information to be used during the at- and post-fieldwork stages, which is provided by Field to File during the pre-fieldwork stage.

6 Conclusions and Future Work

Providing support for activity documentation work is still an open issue that challenges today's computer support efforts. Current applications for activity documentation work are not completely suited to support the documentation of activities and interactions of mobile users in an efficient and complete way. Further, if we consider activity documentation work in RME, the challenge takes on new requirements and restrictions that must be studied and understood in order to determine which would be the features of suitable computer support for this ensemble of activity and environment.

In this paper, we proposed, developed and evaluated Field to File, a system to provide support for activity documentation work in RME. Field to File provides support for the whole lifecycle of activity documentation work in RME by means of mobile and desktop subsystems that allow i) to access the information gathered in a previous stage while working on any later stage, ii) to digitally capture text-notes, pictures and audio files during the at-fieldwork stage, which are automatically transferred between the mobile and the desktop subsystems in a user transparent manner, iii) to automatically name, link and aggregate the captured / generated information resources by the location of the collection site, iv) to directly insert the gathered information resources into a written report in the provided integrated documentation environment.

Evaluation results provide evidence regarding an improvement on the transfer, transcription and insertion time of the captured information when using Field to File. However, it should be noted that the evaluation conducted was an experiment in a controlled simulated setting and with a limited number of participants (10).

Regarding future work, it is mainly directed towards i) an in-situ evaluation at an actual RME and with a significant number of participants (at least 30) performing their actual activity documentation work, and ii) the collaborative construction of written deliverables, which requires explicit support for information sharing, coordination, communication and production among activity documentation workers.

References

1. Hertzum, M.: Six Roles of Documents in Professionals' Work. In: Proc. of ECSCW 1999, Copenhagen, Denmark, pp. 41–60 (September 1999)
2. Morán, A.L., Casillas, R.: Understanding Activity Documentation Work in Remote Mobility Environments. In: Proc. of HCI (10), pp. 1023–1032 (2009)
3. Rogers, Y., et al.: Ambient Wood: Designing New Forms of Digital Augmentation for Learning Outdoors. In: Proc. Intl. Conf. for Interaction Design and Children (IDC 2004), Maryland, USA, pp. 1–9 (April 2004)
4. Sellen, A., Harper, R.: The Myth of the Paperless Office, 1st edn., p. 242. MIT Press, Cambridge (2003)
5. Yeh, R., et al.: ButterflyNet: a mobile capture and access system for field biology research. In: Proc. ACM CHI 2006, Montréal, Québec, Canada, pp. 571–580 (April 2006)

Trends, Challenges and Promises of Rich Experience on Mobile Devices

Yihsiu Chen

UX Manager, Device Marketing and Strategies
Verizon Wireless
yihsiu@acm.org

Abstract. This article summarizes the observation of trends, methods and challenges to provide rich content interaction on mobile devices in the industry, as part of a panel discussion of usability issues related to rich-content applications. The industry and expectation of human-device interaction changes as new wireless and device technologies evolve to provide higher speed, more continuous interaction, and possibility to consume contents while user is on the move. However, the fundamental differences of user needs and ease to deploy applications to fulfill these needs will eventually determine the market between mobile applications and mobile web.

Keywords: Rich content, mobility, mobile web, mobile applications, touch screen.

1 Introduction

In a web development conference recently held in Chicago, instead of the regular paper programs or a CD containing an electronic copy of proceedings, the conference gave each attendee a 16G iPad preloaded with an application specially created for the conference. What was most intriguing to me was not the decision to use an expensive device as a content delivery vehicle – after all, the cost must have factored into the registration fee – but rather the decision to create an application. As the conference is about web development, why wasn't it delivered as a set of web documents led by a bookmark?

1.1 Parallel between One Media to Another

The answers reflect several perception and technological differences in the mobile industry landscape. The obvious reason is the capabilities of the chosen device; iPad has no removable memory like Micro SD card or internal memory that is addressable by browser. However, even without these obstacles, would a bookmark be created instead?

Since the beginning of media and communication, we have seen the business world striving to create a branded presence in the marketplace. After newspapers, yellow pages, radio and TV, companies had been trying to create electronic presences in every platform that was available, including at one time CompuServe and AOL.

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 13–20, 2011.

Many companies were compelled to build a website, without a clear or persuasive connection between the web presence and the brand image of the companies. The consumers have seen many websites filled with great visual effects but very little information or value for the users.

Is the "Apps Store" another platform over which we find another manifestation of such urgency for presence? Where was the mobile web in this rapidly changing trend? What are the right user experience strategies to choose the right medium and design? This short paper does not represent an empirical study; nor does it postulate the usability issues commonly seen in mobile user experience design. It intends to provide observations to reflect on the history of mobile web and its challenges, the current landscape, and the strategies to enrich mobile user experience.

2 Rich Experience on Mobile

2.1 Mobile Web and User Experience Challenge, Circa 2006

Until several technologies breakthroughs carried the mobile user experience across the chasm, the experience of browsing the web on mobile devices or using mobile applications was quite limited. There were many early efforts to design the right user experience on mobile devices, but the barriers toward a fun and useful experience were not the design, but the intertwined restrictions of the technological conditions.

- Limited screen sizes: dial-pad devices typically have screens LCD displays that are less than 2.5" diagonal. The screen size can suffice the need to enough number of lines of menu items, while each line can be as adequately visible, but the screen size does not allow rich content to be properly displayed.
- Limited transmission speed of wireless network: the limit on speed also limits the content that can be transmitted and displayed. Most of the WAP applications maximized the text-based interaction and use graphics sparingly in order to gain decent loading time.
- Navigation limitation of directional keys and soft keys: relying on directional keys and highlight-select interaction, moving around a content-rich page is very labor-intensive.
- Text entry limitation by 3x4 keypads: in spite of the strong motivation from users to adapt to multi-tap methods and predictive texting, entering texts on a 3x4 keypad remains a challenge to many users.
- WAP-based browser on devices with limited processing power is not capable of delivering rich interaction to mobile user.
- Closed environment: the application products on mobile devices were largely controlled by the carriers. Users faced limited choices among user experiences that were not appealing.

These mutually dependent conditions limited both the design of interesting mobile experiences. Furthermore, they prohibit the business and development communities to invest in mobile experiences and to profit from them, both in the innovation of mobile applications, and the development of mobile websites.

2.2 Change of Ecosystems since iPhone, 2007

The appearance iPhone sped up the technology migration of carriers and device manufacturers, and created a fundamental change in the ecosystem. The changes were manifested in just about every aspect that was described as a barrier to good mobile experience in the previous section. Since 2007, many technological progresses converge to augment each other's impact on the mobile experience:

- 3G data transport became fast and prevalent.
- Touch screen interaction liberated uses from linear link-navigation/selection interaction toward a point-and-tap interaction, much closer to GUI point-and-click interaction on desktop.
- Screen sizes grew from less than 3" to approaching 4.5" in 2010.
- Breakthrough of browser capability on mobile devices to leverage faster CPU, as reflected in smoother zooming and panning interaction.
- Improvement of virtual QWERTY keypad that allows quick text entries, assisted by automatic error-corrections, predictive texting, speech dictation, and new interactions like Swype.
- Migration of rich content protocols like Flash to mobile browser
- The new model of application distributions broke the entry barriers and allowed participation to the new business opportunities.

It needs to be noted that although the appearance of iPhone and Safari browser changed the perception of mobile browsing, the original selling point was their abilities to approximate desktop browsing. Although it was not the preferred mobile user experience, it was possible to browse the websites originally designed for PC on an iPhone, while no other mobile phones could do without agony.

Business begins to see the mobile application and mobile web as viable opportunities. What immediately followed these trends was the change of web service strategies to start serving websites tailored toward different mobile devices. For the same URL, depending on the screen size, processing speed and browser capability, a hosting service can return over a dozen websites ranging from the old WAP site for dial-pad devices to more desktop-like rich content site for devices like iPhone or netbooks. At the beginning, this created a challenge to companies who wanted to adopt such strategies, because the server could not always discern the device and browser characteristics to provide the best match. User might have invested in a device with advanced features, hoping to achieve iPhone-like experience, only to be disappointed that the server would still dispense WAP-like content to all non-iPhone mobile devices.

The emerging of Android devices in recent years had simplified the problem. Today, between Safari and WebKit-based browsers, most touch-screen smartphone have comparable capabilities, which makes tiering of web services easier.

3 Mobile Users

Even with these breakthroughs, the mobile experiences are still very different from using the desktop or laptop computer. Many articles and on-line presentations have attempted to outline needs of mobile users even before the technological explosion in

2007. The initial impression that a smart device could mimic the desktop browsing experience was soon corrected. The ever-growing screen size will eventually be in conflict with the ease-to-carry (or "pocketability"), and the screen size and device weight may make thumbing texts a burden. After all, we need to reconcile between the size of touchable objects and the limited real-estate on LCD display.

It also needs to be recognized that while most users believe their devices are always on and always connected, they are not always CONNECTING, or at least one should not make that assumptions without realizing the consequences. Unlike PCs with data connection provided by ISPs to premises, mobile data connections consume both battery power and cost user data usage. While there are many applications that makes background data connections, not monitoring these application behaviors is more likely to cause negative consequences on mobile devices than on PC.

3.1 Mobile User Characteristics

However, as the technologies evolve, we need to take a fresh look at the users on the move, and recognize the fundamental differences in needs, while we evolve the design of user experience into a new paradigm. Mobile users have a few things that distinguish themselves from a user in front of a PC:

Mobile user is not always moving, but is always ready to move. Many people use mobile devices not because there is a particular purpose, but because there is a void of particular activity. Transportation is a common user scenario: I have a few minutes to kill before the train arrives. While taking the train I check the weather or play a quick game. While waiting in line for coffee I check my text messages, and quickly respond to a few of them. While sipping that coffee, I can quickly read an article of New York Times. A mobile user now has ways to fill these little gaps between segments of times, or between locations.

Mobile users are easily interrupted. As the occasions of mobile experience can be short and results of opportunities, users also need to keep their mind on other events that are taking place in their contexts. When the train arrives, I have to take my attention off my phone and board the train; I also need to quit or pause my game when it's time to get off the train.

One should not assume that the mobile contexts allows user to allocate a great deal of attention to the application. This should encourage a strategy to design simpler experiences with careful management of users' attention, instead of complicated applications that overload the experiences with choices and navigations.

Mobile users have impromptu needs of information. Many mobile use cases are built on the assumptions that needs for information emerges frequently and unexpectedly, and that there is a variety of types of information being sought by the mobile user. Over all, these are fairly plausible assumptions. However, many use cases also assume that user welcome information being pushed to them when the opportunities arise, which needs to be validated carefully. Mobile users would like to be informed, but only in the area of their interests, and at the pace with which they are comfortable.

Contexts: time, space and people. Many user experience experts have argued mobile users have different spatial and temporal relationship with their environment. The

"Here and Now" and social relationships are interwoven in popular mobile services we see in the marketplace:

- Space: location-based services; geo-tagging; maps
- Time: taking photo and share, RSS, weather; augmented reality
- People: tweeting, updating your status through social networks and receiving updates from social circle.

3.2 Design Implications

In my own experience, designing for mobile experiences requires a departure from the approach of traditional Information Architecture, and focusing on the of-the-moment tasks. It also requires the designers' discipline to concentrate on the user's attention and intention, rather than peripheral use cases that can be good-to-have.

Simpler structure and reduce cognitive load. For example, the navigation paths from the top level to the bottom should be extremely short. One should refrain from organizing the experiences into navigation charts and rely on sequential steps to achieve the use cases. The design should use GUI treatments and optimal position of touch objects to direct user's attention to engage in functions that are most expected, whereas less important functions are simply discarded. One can also leverage additional gestures, such as double-tap, and drag-and-drop, tilting, panning, and flicking to speed up the executions actions.

Focus on mobile use cases, not consistency with existing website. This sometimes forces a discrepancy between the mobile web design and the design of existing websites, which leads to a concern of inconsistency. However, strong this need for consistency may be, the discrepancy is necessary. As described in the previous section, the mobile users have different needs, and there is no guarantee that the mobile users are interested in the same information or activities as the PC users. Designing for mobile experiences rely on the due diligence of UI design – understand your users in the contexts in which they use the design.

Pay attention to touch comfort. One should recognize the affective impact on the user experiences, especially in the following aspects of GUI design: using nicely rendered GUI elements to create saliency to manage user attention and reduce obstacles to a smooth interaction. The obstacles can come from (a) difficulty to identify plausible action choices, if there are too many equally plausible candidates, (b) low physical comfort to attain the choice, if the target is too small to touch, and (c) misplacement of important touch objects and make them ergonomically difficult to touch or forcing user to use two hands. These obstacles can create a higher anxiety level that failed to be captured by traditional usability testing methods, but can influence the subjective pleasure of using the device.

4 Mobile Web or Mobile Apps?

When it comes to delivering rich content to mobile user experience, mobile applications today remain the more viable venue. However, the challenge from mobile web technologies, especially HTML5, presents a real threat the dominance of mobile applications.

To understand the strength and weakness of these approaches and to find the right strategy to deliver mobile experiences, we can start from a simple comparison:

Table 1. Comparison between mobile apps and mobile web

Mobile Application	Mobile Websites
Strength: - More mature and rich-content ready - Established distribution channel and revenue model - Ability to integrate with native capabilities, such as telephony, SMS, sensors, and camera	Strength: - Updated quickly on server; no need to push out changes to numerous clients - Wider addressable market; easier to port across device platform with comparable abilities
Weakness: - Harder to update - Harder to port across device platforms	Weakness: - Less mature in delivering rich content - Monetizing opportunity is less clear - No integration with native capabilities.

The table of comparison may be clear and logical, but such comparison already existed between PC applications and websites serving PC browsers. What implications on mobile experiences can we draw from these contrasts?

Table 2. Most downloaded applications on from Apps Store

Free apps		Paid apps
- Crash bandicoot	- Pocket God	- Facebook
- Koi Pond	- Cro-mag Rally	- Pandora
- Enigmo	- Ocarina	- Google Mobile App
- Bejeweled	- FieldRunners	- Shazam
- iBeer	- iFart Mobile	- Movies by Flixster
- Moto Chaser	- Touchgrind	- The Weather Channel
- Pocket Guitar	- iHunt	- Google Earth
- Flick Fishing	- iShoot	- Bump
- Tetris	- Monopoly	- Skype
- Texas Hold'em		- Paper Toss
- Super Monkey Ball		

We can employ data from a different source to assist in our understanding of the difference and opportunity of displacement, by examining the applications that are already popular in the marketplace. In Table 2, the top downloaded applications are listed, based on Apples' own records of downloads since July 2008. (http://itunes.apple.com/WebObjects/MZStore.woa/wa/viewCustomPage?name=page1BillionApps).

Based on the list, we can divide the apps into two large categories, knowing that each category represent hundreds or thousands applications similar to the ones represented here.

Type A: Information, always in touch. Candidates for this category include news reader applications affiliated to different media (e.g., New York Times, CNN, and Wall Street Journal), iBook, and Weather. They also include social networking applications like Facebook, or information apps such as the ones providing interface into Wikipedia. These applications share the common traits that they are filled with information. They also rely on connecting to sources on-line to present instantly updated content. Some of these apps (such as ESPN) distribute video contents, but for most of these apps the contents constantly updated but are static.

Mobile apps do not have a clear advantage over mobile web when it comes to Type A applications. The characteristics of Type A applications are similar to that of mobile web, and most of these apps are free to download and to use. The apps do not monetize solely by distributing the apps themselves, so there is no disadvantage to deploy the same content over mobile web. In fact, in some cases the look and feel of mobile apps and mobile web can almost be interchangeable within the Type A applications.

Type B: Communication, entertainment, media, and leave-me-alone. One would immediately notice that games populate a large share of the free application category. Games, accompanied by communication, media consumption, and GPS-assisted apps such as Google Mobile application, make up most of Type B category.

Applications in this category have higher requirement for rich-content interaction. Users are more involved with the manual interaction with the device, and expect satisfaction of emotional and entertain needs rather than information needs. They also expect these, especially games, to work off-line.

The established distribution model of these apps is essential to their monetization, which presents a challenge to mobile web. In summary, these apps have higher requirements of off-line usage, more intense interaction and integration with device capabilities, which make them more resistant to being displaced by mobile web.

5 Conclusion

The success of rich-content user experience on mobile devices depended on the maturity of technological enablers, which came into reality only in the last few years but since then has progressed at a lightning speed. The evolution continues, and we are likely to see the following enablers breaking grounds in the next two years:

- High-speed 4G networks being widely deployed starting 2011
- New battery technologies that can sustain longer use time
- Super AMOLED screens that are thinner and with more clarity
- Dual-core devices with high processing power
- Ability to capture and consume 3D contents on mobile devices
- Maturity of HTML5 in competition with other standards.

These enablers will challenge the designers' ability to envision innovative user experiences, and at the same time present new usability problems such as touch-screen accuracy and comfort to interact with 3D displays. In spite of these new technical trends, the core users' needs are like to remain consistent, and the objectives of the mobile UX designers continue to be simplifying the user model of mobile services, and making the mobile applications richer yet more transparent to the user.

Finding Suitable Candidates: The Design of a Mobile Volunteering Matching System

Wei-Chia Chen[1], Yun-Maw Cheng[1], Frode Eika Sandnes[2], and Chao-Lung Lee[1]

[1] Department of Computer Science and Engineering, Tatung University,
No. 40, Sec. 3, Zhongshan N. Rd. Zhongshan District, Taipei City, 104, Taiwan
[2] Department of Computer Science, Oslo University College
P.O. Box 4, St. Olavs plass, N-0130 Oslo, Norway
ann770402@gmail.com, kevin@ttu.edu.tw,
frodes@hio.no, d9806006@ms.ttu.edu.tw

Abstract. It can be difficult to get started with voluntary work for potential volunteers (PVs). Moreover, it is difficult to find and recruit suitable candidates for nonprofit organizations. To help solve this problem we designed a mobile matching prototype that enables an organization to actively promote ongoing volunteer activities with the need of recruitment through their bubble icons on an instant map. In the other end, PVs can easily get started by monitoring the colors of the icons and tap the ones which match their interests. This allows them to read about developing threads and browse the corresponding activities. The system is evaluated by interviewing two organization managers and three volunteers.

Keywords: Mobile Device, Communityware, Volunteering Matching, Context-awareness.

1 Introduction

Voluntary work is making an important contribution to society, for instance, supporting social services, sporting events, religious activities, etc. However, it is difficult to become a volunteer and maintain the commitment [6]. A national study reports that 57% of organizations in Canada have difficulty recruiting the type of volunteers they need and 49% have difficulty retaining volunteers [6]. Thus, it is hard for potential volunteers (PVs) to get started, it is hard for organizations to find suitable candidates and it is hard to maintain the commitment.

Mobile devices are recognized as a helpful tool for establishing contact between PVs and organizations. Moreover, context-aware attributes such as the users' location, time, and current activity can be valuable attributes for assessing relevance [8][15][10]. To recruit the right volunteers for a task, a suitable means of contact should be established between the PVs and the organization. The meaning of "suitable PVs" is tied to friendship, social integration, characteristics of the job and the PVs commitment to remain with the organization over time [3]. Hence, "suitable" should not only refer to the reasons for joining, but also extend to future working partners

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 21–29, 2011.

and environments. Consequently, we believe a mobile matching system can bridge the gap between PVs and organizations.

In order to understand how to reduce the sense of distance, we investigated the existing matching practices. Moreover, to establish the practices for finding the suitable candidate, including both sides' selection criteria and the meaning of “suitable,” we surveyed the literature and interviewed volunteers and organization managers about what is important for them before and after they joined the organization. The investigation concluded in five design decisions that are the foundation of the mobile matching system prototype described herein.

2 Related Work

A new form of engaging volunteering have recently emerged, namely micro-volunteering or mobile volunteering, which makes volunteer work more accessible by the means of mobile technologies. Some organizations have started to crowdsource simple tasks online. These tasks, accessible through smartphone applications, only took participants a few minutes to complete. Examples include photo-tagging wildlife, mapping green urban spaces, making an audio map for the blind or filling in a research survey, etc. Although smartphones can help promote volunteering wider and making it easier to start, organizations are skeptical of micro-volunteering [15][12]. Most volunteers still follow traditional patterns as reported by three national surveys [4][14][13].

Conventionally, organizations publicize opportunities via posters, newspaper advertisements, public appeals on TV, radio, and the Internet. This approach only allows organizations to reach PVs passively. PVs who get involved in this way contribute more hours than those who are asked to volunteer. Organizations reach more people than just relying on their social network. The other approach is to actively recruit PVs based on the social network of organizations’ fellow members. The advantage is that organizations can get feedback about who are interested in a more trusted manner. PVs can have lower motivation requirements and more familiarity with organizations because insiders can help convey information about the organization [3].

As mentioned above, PVs need a way to investigate the organization before they make a decision about volunteering to help bridge the gap between reality and the PVs’ expectations. Jumo[1] gathers latest news and talks of the projects that a user subscribes on one’s homepage. The user can communicate to them on the page of the projects. Likewise, doGooder [11] shows others’ testimonials to guide and motivate PVs. Besides, it is easier for PVs to feel confident when a commitment is based on assessing several options. Volunteermatch[2] and niceSERVE[3] are location-based smartphones applications that allow users to find opportunities nearby while the second focuses on short and trial events to help start. Several studies show that the chance for someone to volunteer is almost three times higher if the person is asked to volunteer than if not asked [3]. Tzu Chi[4], an international humanitarian organization,

[1] Jumo, http://www.jumo.com

[2] Volunteermatch, http://www.volunteermatch.org

[3] niceSERVE, http://www.niceserve.org

[4] Buddhist Compassion Relief Tzu Chi Foundation, http://tw.tzuchi.org

has used a similar method, known as concentric circle recruitment, to successfully extend its number of volunteers.

The implications herein are that the mobile matching system should adopt a simple way to let PVs get accurate picture and enough information about the opportunity they are interested in. Also, it should let organizations be actively aware of suitable PVs and invite them to join, no matter whether they know insiders or not.

3 Finding Suitable Candidates

We not only want to reap the benefits of current matching practices but also eliminate their shortcomings. A common problem with the two methods for organizations is that they know very little about whom the PVs are and what they really want. On the other hand PV are unable to compare people, organizational culture and working environment inside of an organization with another and are consequently less able to make reflected decisions about volunteering.

Therefore, we need to establish what factors PVs and organizations look for in counterparts. We combined the opinions we collected from a focus group interview as well as related research. We interviewed two participants, an organization manager and professional volunteer from Eden social welfare foundation, both female. They have both volunteered for more than two years, and the manager has experience from many organizations in different areas, such as education, social service and religion. The organization they are serving is one of the national-wide nonprofit organizations in Taiwan, mainly providing social services for people with disabilities. We also surveyed previous studies about the need of organizations and volunteers in various areas to get a more complete picture.

For organizations, the ultimate goal of recruiting is to find candidates who can help accomplish the organizations' missions. Thus, the requirements vary across organizations. One trend however is that they all prefer candidates who can provide long-term commitment [5]. For PVs, primary considerations include task time, skills needed, task location, and people they know are already involved [8]. The reasons for staying are very different. After joining an organization, friendships, social integration in the workplace and working environment are primary determinants of whether to stay [3][2].

Three implications emerged from above investigation:

Providing context of volunteer opportunity. Time, skill needed and location are three basic criteria relevant to PVs. In addition, if the system is allowed to make use of the user's social network account to show activity members they know then this can also help find suitable candidates since one of reasons to volunteer is to meet friends [4].

Improving team social integration. Team social integration is a multifaceted construct including elements of cohesiveness, satisfaction with coworkers, positive social interaction, and enjoyment of team experiences. High team performance pursued by organizations is positively related to team social integration [7]. PVs and organizations should be matched such that PVs have good chances of successfully integrating with the people already working in the organization. Several studies on organization behavior indicate that people who have similar psychological characteristics, including personality, attitudes, and values have better social integration, task performance and satisfaction than those who are more dissimilar. Volunteers are then more willing to stay and the organizations perform better [7].

Understanding the motivation. The need for a similar or interesting organizational cause is one of motivations to volunteers. When PVs whose motivational concerns are served by their opportunities or organizations gain relevant benefits, PVs feel more satisfied, leading to stay [1]. The matching system should therefore find which functions are important to the PV and match those with the task or people working for the task with the same function.

4 System Prototype

This section presents the mobile matching system prototype based on our findings. It is an instant volunteering matching system that can recommend PVs the most suitable ongoing opportunities. This design aims to solve the difficulty getting started and finding suitable matches between individual PVs and organizations.

4.1 Scenario for PVs

Bob, a potential volunteer, wants to make a positive contribution. He discovers this system and would like to try. At the beginning, a map shows up containing all ongoing volunteer opportunities around him according to his GPS location. Meanwhile, a

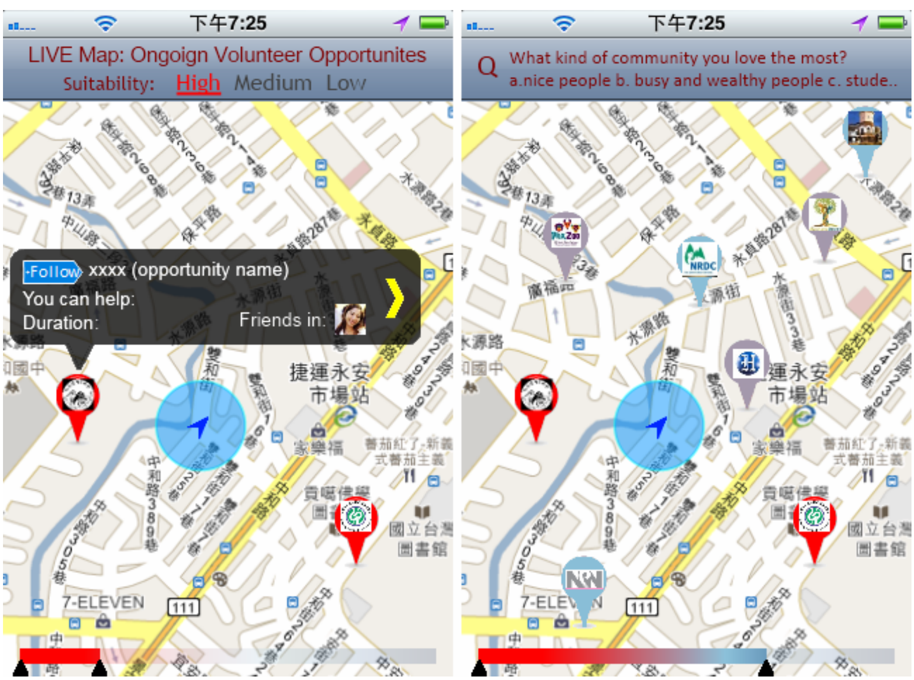


Fig. 1. *Left*: the system screenshot after opportunities are personalized. *Right*: the screenshot for wider range of suitability.

notification from the system asks if he needs the opportunities ranked according to his preferences. Bob taps on the “yes” button because he feels there are too many volunteer options. Thus, Bob is navigated to a new page which tells him he is going to answer three quizzes for better recommendations. When done, the map is shown again with fewer opportunities displayed (see Fig. 1 left). In the middle, the blue arrow is his current location. Red icons are opportunities ranked the most suitable. He can adjust the suitability bar or zoom out the map to see more options (see Fig. 1 right). Bob sees the organization logo and end time of the opportunity, skills needed, mission goal and his friend retrieved via his social networking account. Bob can then tap the yellow arrow at the right to get more information.

Asking. Anna, Bob’s friend as well as a volunteer, is notified by the system that her recruitment of activity matches Bob’s preference when he is currently finding a volunteering opportunity. Anna is asked if she would like to invite Bob to come and take a look. When she taps “yes,” an invitation from Anna is sent to Bob, and at the same time, the invitation from a person belonging to the other opportunity with high suitability is sent to him. Bob feels more confident about these opportunities and considers taking a look. This inviting mechanism not only strengthens the ease and accuracy to ask people to volunteer, but also improve non-volunteers’ barrier of not being asked [4].

Watching online. Suppose that Bob does not go immediately but would like to “save” the two opportunities and others he found interesting. He taps the blue “follow” tags to monitor the progress of these activities. Hence, the organizations that post the followed activities know Bob is a potential candidate from a follower list, so they can share information initiatively relevant to what he cares about on the activity page instead of traditional one-way posts. This type of interaction can also apply to the situation if Bob goes take a look or participates, because “follow” can let him and activity members keep in touch to improve retention.

Instant participation and watching offline. Suppose Bob accepts Anna’s invitation, after he evaluates the ending time, and for a while, he arrives at the venue of the activity. He is provided the other way to “watch” the opportunity. He can immediately join and help if he wishes and is allowed. This way to approach the opportunity makes Bob feel that becoming a volunteer is not difficult. Besides, the whole process gives Bob a real and vivid image about the opportunity, volunteer partners, and the organization. It not only helps him decide or strengthen the motivation but also gives the activity members a chance to informally interview his suitability before joining formally. Briefly, PVs can interact with volunteer opportunities at an appropriate moment. In addition, they can have better control over their time committed to the volunteer activity.

4.2 Scenario for Organization Members

To recruit PVs, Anna and her activity members need to post the icon of their volunteer activity on the map of the matching system. First, they tag the activity on their accounts created after answering quizzes, and then just check their locations to tell the system where the activity is happening. They are not worried about privacy problems because it is the activity that shows on the map. They can share instant situation on the activity page. When Anna allows her friends to see what activity she is participating in to further spread the news, she is notified Bob is interested. Clearly, spreading the news of volunteer opportunities and promoting the progress of the activity can be done by everyone.

4.3 Matching Mechanism

The ranking of opportunities for a person is according to similarity in personality, attitude, value and interest between activity members and the user. The similarity value comes from comparing their answers to a quiz. People who have the same or similar answer to questions are defined as suitable [9]. Multiple questions are packaged as a fun quiz or a single question is a quiz. Users get a result after finishing each quiz. For example, if the title of a quiz is “Which EU capital should you come from?” “What career is right for you?” or “Understanding your interpersonal relationships from smartphone?” Although the quizzes seem to be irrelevant to volunteering, the answers are the key to build a profile of a user.

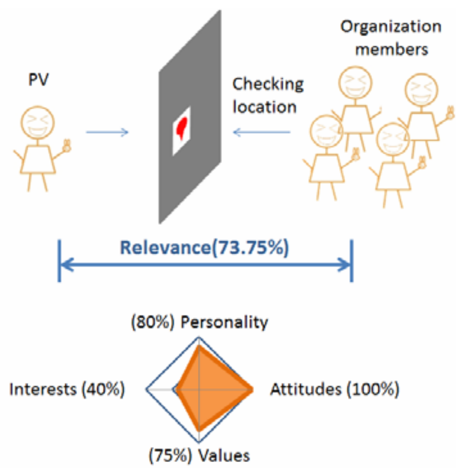


Fig. 2. An example of matching mechanism

Taking the one-question quiz “Understanding your interpersonal relationships from smartphone” as an example:

Q: Why do you use your smart phone?

- Convenience
- Necessary for working
- Fashion, so I can show off
- A gift given by others
- Cute and special, so I have several of them
- For safety concern

The user's result is a description of interpersonal relationships according to the answer. Then one can either "self-correct" the result by choosing a more fitting description, or go to the next quiz. It makes sure the answer is representative. In the multi-question quiz, answers are representative without the aid of a description, such as "What career is right for you?" quiz:

Q: Do you like to help people?

- Sometimes I like animals more than people.
- I'm good at emotional situations. I'm really in touch with mine.
- I like to give people advice and listen to their problems.
- I like to help them when I'm on their side.
- I love to help anyone in need.

We observe how people play with these types of quizzes on Facebook, and we found that people are willing to answer quizzes because they are fun, people can get feedback about themselves as a reward, or their friends who took the quiz. People take the quiz actively and try to answer correctly to get an accurate result. Another important finding is that these answers and results are not very private so there is strong likelihood of sharing the results with everybody. Thus, organizations can create their own fun quizzes to find out who have the characteristics they want and issue an invitation. Moreover, organizations can understand the motivation of the PV and their own volunteers to improve the work arrangement. Similarly, PVs can find suitable organizations.

Another interesting type of quiz on Facebook is that people answer questions about friends, such as, "Do you think Bob is altruistic?" Our system can give a clearer image to everybody about how an organization and its volunteers may be. All in all, answers to quiz can give a good matching of psychological attributes, which leads to good social integration, and better understanding of motivation.

5 System Evaluation

Five participants including three males and two females, ranging in age from 32 to 55 were recruited. They had experiences of using computers and that of volunteering for more than two years.

We used two different evaluation approaches. Three participants were simply asked to use the system and state their thoughts about the interface and how their organizations can adopt the system. The other two had an adequate discussion on the problems of recruiting and retaining for PVs and organizations before starting the evaluation. Therefore, they provided more insightful perspectives during the evaluation.

The system prototype is made with PowerPoint and shown on a laptop during the interviews. The file contains introduction, quizzes, results and the system snapshots (see Fig. 1.). We began with a short oral introduction about the function of the system and explained how to interact with the slides to reduce unnecessary interruption. When they reached the map stage, the interview started. We got feedback about the interface usability, comments about the system and suggestions of new functions and usage.

Interface. All participants understood that red represents high suitability, but one participant suggested that the organization logos should be replaced by something related to the opportunities, for example the type of volunteer they need. Some participants thought there is a navigating function according to their past using experience of Google map. Another asked for a search bar because it is perceived as a volunteer opportunity searching system. A website version for using at home was required.

Functionality. Most participants considered our system very helpful. "This is a very classy and considerate matching system." They think the number of quizzes was appropriate and all agree with answers not being private. A participant described a scenario he had before and it was exactly the same as what we expected to happen. "A friend asked me if I'm interested in something, so I engaged. I thought it was nice so I remained." He also told us that "If it (the one he is interested in) is nearby, I will go and take a look, but I won't tell them. If it's quite nice, I will hope they can fill me in when they have similar activities." Two participants thought that the established organizations with many activities in fixed places might not need this system. This implied that a different approach is needed for voluntary work at fixed locations.

New functions and usage. A large chunk of the discussions focused on the way to show the suitable candidates in terms of time. One suggested using a scrolling text to show the latest news or opportunity. Most participants indicated they used to get the date of activity one or two weeks ago, so that they can arrange their time. It meant a time bar or a calendar is needed for volunteers to know activities of their organizations in advance. "I want to volunteer on Thursday, so I want to find the opportunity on that day." "I want to find the opportunity at the time I' m free." A participant pointed out the time bar can be applicable as a criteria filter to PVs. In addition, a type filter was suggested for searching, and this reflected the importance to show the logo as the type of volunteer activity. On the other hand, they still valued the feature of instant participating. "They (volunteers) don't have a stable time commitment. The instance can let them volunteer when they are free." "If there is a severe disaster, emergent rescue is needed, or some activities with high mobility," like cleaning the street, visiting clients' homes or education service. One said this could be used to promote some organizations' entertaining activities that are open to everyone.

6 Conclusion and Future Work

We presented an instant location based and mutual people-oriented volunteer matching system for organizations and PVs. It aims to provide an easy way to start by three functions, namely, "invitation," "instance" and "follow" and offer ranked opportunities according to a psychology-based similarity. Participants were positive about the system and suggested several functions and uses, such as time bar and emergent or big events.

Future work includes studying how to bridge micro-volunteering and macro-volunteering. Second, we hope we can integrate people in need into the system so that further and more people can help directly. Ushahidi[5] shows a good example regarding

[5] Ushahidi, `http://www.ushahidi.com`

the importance of mapping the far aids onto locations or people in need when considerable emergent events happened. Finally, our matching mechanism may help people with the same cause find each other to form a group doing what they cannot do on their own.

References

1. Clary, E.G., Snyder, M.: The Motivations to Volunteer: Theoretical and Practical Considerations. Current Directions in Psychological Science 8(5), 156–159 (1999)
2. FormosaSoft Corporation: College Graduates Attend Nonprofit Organizations: the Evaluation Survey of Performance and Effect of Internship. Qualitative Report, National Youth Commission, Executive Yuan, Taipei (2010) (in Chinese)
3. Herman, R.D., Associates: The Jossey-Bass Handbook of Nonprofit Leadership and Management, 2nd edn., San Francisco (2005)
4. Hall, M., Lasby, D., Ayer, S., Gibbons, W.D.: Caring Canadians, Involved Canadians: Highlights from the 2007 Canada Survey of Giving, Volunteering and Participating. Statistics Report, Statistics Canada, Ottawa (2009)
5. Hall, M.H., Andrukow, A., Barr, C., Brock, K., Wit, M., Embuldeniya, D., Jolin, L., Lasby, D., Lévesque, B., Malinsky, E., Stowe, S., Vaillancourt, Y.: The Capacity to Serve: A Qualitative Study of the Challenges Facing Canada's Nonprofit and Voluntary Organizations. Canadian Centre for Philanthropy (2003)
6. Hall, M.H., Wit, M.L., Lasby, D., McIver, D., Evers, T., Johnston, C., McAuley, J., Scott, K., Cucumel, G., Jolin, L., Nicol, R., Berdahl, L., Roach, R., Davies, I., Rowe, P., Frankel, S., Brock, K., Murray, V.: Cornerstones of Community: Highlights of the National Survey of Nonprofit and Voluntary Organizations, Statistics Canada, Ottawa, ON (2005)
7. Harrison, D.A., Price, K.H., Gavin, J.H., Florey, A.T.: Time, Teams, and Task Performance: Changing Effects of Surface- and Deep-Level Diversity on Group Functioning. The Academy of Management Journal 45(5), 1029–1045 (2002)
8. Kane, S.K., Klasnja, P.V.: Supporting Volunteer Activities with Mobile Social Software. In: Ext. Abstracts CHI 2009, Boston, MA, pp. 4567–4572 (2009)
9. Knippenberg, D., Schippers, M.C.: Work Group Diversity. Annu. Rev. Psychol. 58, 515–541 (2007)
10. Loke, S.: Context-Aware Pervasive Systems, New York (2007)
11. Morse, J., Cerretani, J., Halai, S., Laing, J., Perez, M.: doGooder: Fostering Volunteer Communities to Serve the Homeless. In: CHI 2008, Florence, pp. 3843–3848 (2008)
12. Ward-Smith, J.: Making Minutes Matter – Orange Mobile Micro Volunteering, `http://www.i-volunteer.org.uk/jamiet/many-minutes-matter-orange-mobile-micro-volunteering`
13. Taiwan Directorate General of Budget Accounting and Statistics: Channels by Which to Perform Voluntary Service in the Recent Year, `http://win.dgbas.gov.tw/dgbas03/ca/eng_social/XLS/92t3-2.pdf`
14. U.S. Bureau of Labor Statistics: Volunteering in the United States—2009. Statistics Report (2010)
15. Zak, D.: Mobile Volunteerism, `http://www.dalezak.ca/2010/06/mobile-volunteerism.html`

The Effort of Social Networking on Social Behavior – Integrating Twitter, Mobile Devices, and Wearable Clothing as an Example

Chen-Wei Chiang and Kiyoshi Tomimatsu

Kyushu University
4-9-1 Shiobaru Minami-ku Fukuoka 815-0032, Japan
interaction.tw@gmail.com, tomimatu@design.kyushu-u.ac.jp

Abstract. In this research, we proposed a new communication method system which is based on Social Network Services (SNS). As the medium, we used a T-shirt with an AR marker printed on it (AR T-shirt). User wears the AR T-shirt and has their Tweets and latest messages shown on a mobile device. Typically, SNS users exchange personal messages or information in virtual world, web site or mobile interface. Therefore in this research, we tried to establish a new system to explore users behavior. In this system we utilizes Twitter, AR T-shirt, and mobile device (iPhone) to propose a new communication method. When people use iPhone to recognize AR marker on the AR T-shirt to communicate, it will initiate a new communication method. Users can directly send and receive the latest personal messages. Thus, this system will transform virtual communication into the real world.

Keywords: Social network service, wearable clothing, mobile device.

1 Introduction

Social Network Service (SNS) is currently affecting internet user behaviors. Users are getting accustomed to interacting with their friends via the internet. As people use SNS more and adjust their internet usage behavior, their social behavior will be affected as well. In this paper, we proposed a new system as a tool for the user to interact with the social media. Our proposed system combines Twitter, marker based AR and a mobile device. By using this system, we discovered a new method of interacting via SNS in the real environment. On the other hand, we can also use this feature to develop other SNS applications or contents.

2 Background

As people use SNS more and adjust their internet usage behavior, their behavior in the social environment will be affected as well. Nowadays, exchanging phone numbers has developed into exchanging email addresses, instant messenger, Facebook, and Twitter accounts. People have multiple ways to communicate now, as more choices of media are available today [1].

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 30–37, 2011.

SNS is changing internet user behaviors. Users are getting accustomed to interacting with friends on the internet. Social network services have become more diverse; for example, users are able to publish news about themselves to others in the same network through different media at any location and time. One of the most common forms of SNS media is text-based message. Within the same network group, users are able to share their text-based news to each other instantly.

For example in recent Japanese culture, most internet users use Twitter as their main information exchange platform. With the increasing popularity of the iPhone and other smart phones, the utilization of Twitter and the ways of expressing their tweets extend beyond the scope of internet. Users are able to talk about what is popular on Twitter when they meet. Because of this, more people who are using Twitter are able to follow the current trends.

Furthermore, we can observe that people's behaviors may change when they are following or being followed by someone. Their relationship may become closer, and they can continue to communicate their tweet contents in the real environment. This interacting pattern will likely to change their relationship.

2.1 Augmented Reality Technologies

In this paper we utilize augmented reality technologies as the main tool to provide a new mobile interaction. Augmented Reality (AR) can be defined as real-time direct or indirect view of a physical real world environment that has been enhanced / augmented by adding virtual computer generated information to it [1].

Augmented reality technologies began in the 1950s and had underwent a thorough development until present. Nowadays, along with the new advances in technology, the amount of AR systems and applications has been produced, most notably smart phone and PDA applications. The details of Augmented Reality Technologies will be discussed below:

Computer Vision Method in AR: In augmented reality technologies, computer vision usually consists of tracking and reconstructing. The tracking process consists of detecting fiducial markers or optical images. Augmented reality technology later reconstructs the data obtained to reconstruct a coordinate system.

AR Displays: Although there are several types of AR displays, in this paper we are limiting the AR display scope to handheld displays. Handheld device mainly consist of a small processor with a display. However, recently a portable device that provides a combination of a powerful CPU, camera, accelerometer, GPS, and other sensors that is known as the smart phone has also been developed, making it as one of the most suitable device to provide AR interactions.

Input Devices: The input devices for AR is greatly depend upon the type of the system. In this research, we used a mobile phone as the handheld device as the input device.

Tracking: The most commonly used tracking device in augmented reality is digital camera. However, due to the development of the technology, other devices have also been developed to identify AR markers. Recently, the built-in digital camera smart phone has the capability to recognize images in high resolution and could be used as an alternative to track AR markers.

2.2 Related Work

In this section, we will discuss some works that are related to our proposed system. The concept about connecting SNS into the real world environment has been proposed in the past, such as Panasonic Video T-Shirt, Department Store Guidance, and various iPhone Applications. Each of these works uses different media and communication method. The details of each of these works will be discussed below:

QR Stuff: Print QR Code Service. In this service [3], users can input the text or contents by using web site interface. After inputting the contents, this web site can help users to generate a QR code according to their input. Alternatively, users can print the QR code on T-shirt, cups, mugs, bags, etc. This service enables people to dynamically create their own custom QR code products while sharing information effectively.

AR Greeting: Panasonic Video T-Shirt. AR Greeting: Panasonic Video T-Shirt [4] was made to promote Panasonic's latest television model which is called Viera. It was designed as a promotional event. Using this service, they used marker based AR for displaying animation. The AR marker was printed on T-shirt and users can use iPhone application to recognize the AR marker and see an animation via their iPhone.

In this work, users can experience an interesting interaction experience and they can pass the experience to their friends further. When they pass their experience to their friends the numbers of the user will build up. Thus, this behavior can be very effective as a promotional tool.

AR Tees: AR Code T-shirt. Certain information that is recorded on the QR code can be printed on a T-shirt[5]. Thus when the QR code is scanned, corresponding information about the user can be provided, such as personal contact, etc. Recently, the AR Tees contents are developing into multimedia content, such as video streaming. This QR code and augmented reality technology would likely to form a new interactive mode of communication.

N-Building: Department Store Guidance. QR code is used for floors guidance in department store. Users can check the available information of each floor by using an iPhone application [6]. The screen can show real time information, and the content can also be changed depending on the current promotional event of the department store. In this work, we can also find that after recognizing the QR code, the screen can present a real time animation as the guidance. However, a smart phone is necessary to set this augmented reality environment.

iPhone Applications. There are many iPhone applications which are designed to access Facebook, Twitter, and other social network services. Most of these applications are designed to read and post messages directly as their users travel. Meanwhile, some applications for exchanging personal information are also available. In general, most of these applications are designed to access available features in social network services while providing mobility to the users.

QR Code Application. In the past, the main function of mobile phones is to provide auditory communication. Nowadays, people can also use mobile phones to take photos, send E-mails, or browse web site. These new smart phone features have changed user behaviors. In particular, phones that have built-in QR code recognition functionality are very common in Japan nowadays. This QR code technology helps people to receive information easily, efficiently, and more conveniently. For instance, when scanning in a QR code from an advertising poster, consumers can receive information about the store or promotions straight into their cell phone instead of input it manually into their phones and do a search. However in this context, QR code only acts as a medium for one-way information transmission. In this research, we focuses on developing the QR code possibility as the medium for two-way information transmission.

Twitter: Social Network Services. Twitter is a social network service. People can exchange or post their daily life sentence on twitter. Twitter is a web based social network service [1]. Recently, the demand on smart phone users is increasing. Social network service also offers their service on smart phone. Thus, users almost can share everything by using smart phone. And people can be much closer via this service.

2.3 Research Framework

Based on the user experiences of using mobile phone to receive information via QR code, the system is established with the following three conditions:

1. We consider the mobile phone with a built-in camera as the appropriate platform in this research. When presenting diverse information, QR code is still limited. Thus in this research, we propose the Augmented Reality (AR) technology to provide real-time responses and two-way communication.
2. Users usually find QR code to be related to certain information, which provides an affordance for the QR code itself. If they want to receive the information, they can use their mobile phones to recognize QR code. Based on this user experience, we proposed to AR markers on clothing.
3. This research is under the hypothesis that when users find AR markers on someone's clothes, they can find the information of the person through AR markers. Thus, this research combines AR technology with existing SNS services by using mobile phones to recognize AR markers on clothing.

In this research, we utilized 3 main frameworks, i.e.: QR Code that acts as the medium; AR system which is displayed when the user scan the QR Code; and a smart mobile phone to recognize the QR Code thus providing user interaction, along with the 3G network provided by the network provider to transmit and receive the corresponding data.

3 System Design

In the previous section we have discussed our proposed concept of combining mobile device, SNS, and QR code to design a new communicating method. In this section, we will discuss possible communicating scenarios and the system design in order to make the system easier for the users to comprehend and use.

To be precise, we designed a communicating system which consists of an AR marker and a T-shirt as the tangible medium. This system is designed to recognize marker based AR system which is printed on the T-shirt. Therefore, users can communicate directly and read latest message by placing their smart phone into the AR marker and having the QR code recognized.

3.1 Scenarios of Communication Behaviors

For establish this system, predicting users' behaviors via scenarios can help us to get an overall view of this system. Specifically, users wears a T-shirt with an AR marker printed on it (AR T-shirt) and uses handheld devices to interact each other. This research focuses on three situations to start the two-way user communication scenarios: (1) Both user wear the AR T-shirt and use mobile phones to exchange information to the other person. (2) Only one user wear the AR T-shirt and the other one uses a mobile phone. (3) One user wears the AR T-shirt but this user does not know that the other user is using a mobile phone to receive information.

3.2 System Design

Based on the scenarios and requirements, the system is divided into three parts:

1. Clothing with AR markers: In some circumstances, one of the users may stand behind of the other user. Thus, AR markers are printed on both side of the clothes, the front and the back. This design allows users to take mobile devices to read positive or negative information from the AR marker.
2. The mobile phone application to recognize AR markers: After the user opens the application, the AR marker identification is executed and the corresponding information will be displayed on the mobile phone screen.
3. The mobile phone application for sending and receiving SNS information in real time.

3.3 System Architecture

Our system was developed on iOS for iPhone. In our proposed system, we used the iPhone to recognize AR marker. IPhone has a built-in camera which is able to capture and present the result in the real-time. Furthermore, to recognize AR marker, we use AR Toolkit, i.e.: a library which is designed for iOS to build Augmented Reality (AR) applications. By using AR Toolkit, we can create a marker based AR recognizing system quickly and effectively.

Furthermore, considering that internet connection plays an important role in our proposed system, we utilized iPhone's 3G signal to send and receive messages (Tweets). By utilizing these platforms, we believe that our concept was sufficiently represented on this system.

3.4 Installed Features / Services

As discussed in the previous section, we propose the design to be a real-time interactive SNS system. The system is designed to allow user to access the SNS via AR system in real-time, thus initiates dialogue and interaction between users. Therefore, considering those needs, three main features were developed, i.e.: interactive mode, feed, and user interface.

Interactive Mode. As mentioned in the interaction scenarios (3.1), there are two main roles for the basic interaction mode, i.e.:

1. The first user who wears the AR T-shirt;
2. The second user with the mobile phone who acts as the observer.
3. The interaction is carried on as the second user scans the QR-code on the first user's AR T-shirt and initiates two-way communication.

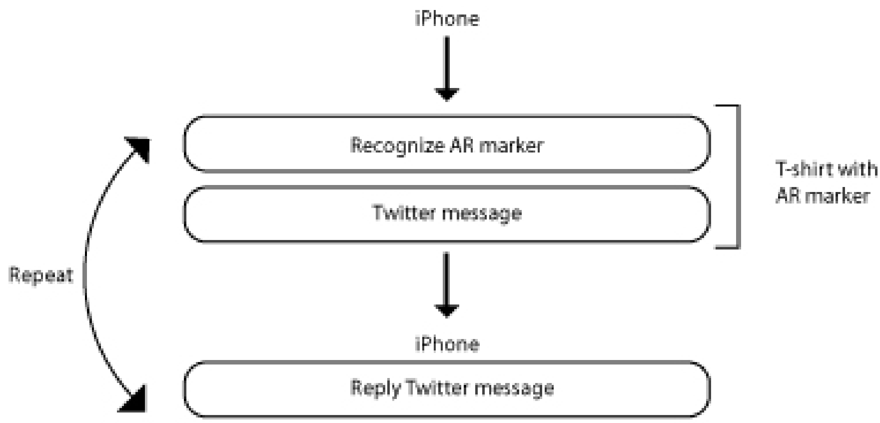


Fig. 1. Interaction Mode

Feed. In this system, we provide the user the information through the identification AR to get the latest twitter feed. Twitter displays the feeds in a timeline and sorts them automatically, displaying the latest feed on top. Therefore, in order to provide the most relevant information, we selected the latest twitter feed to be displayed on the screen. This method allows user know that the information is being constantly updated. In addition, when the user touches the latest feed, they will be redirected to a page with the complete twitter timeline.

User Interface. We used the original iOS toolbar setting in order to provide the user a familiar user interface. In addition, we also provide visual feedback to let the user know their action after touches the screen menu. The original iOS toolbar for iPhone user is much easier for using a new interface. Users are already have iOS operate experience. When they face a new application, they can learn to how to use the new interface fast.

3.5 Layout

In this system, the most important user interface is iPhone application interface. Thus, We designed it in a simple way. Users enter this application and they can directly use a real-time camera capture function to recognize AR marker. After recognized, users can check the latest twitter message via the screen. If users want to reply the message, they can touch the screen directly and it will switch to reply window for inputting text.

Fig. 2. The interfaces to recognize AR code in iPhone

Users also can directly push reply button for replying message and go back to the first screen to check the latest message.

Considering about iPhone users may stand in different place for recognize AR marker. The AR marker is printed on two sides of T-shirt. Depends in this system, we can study users how to use this system and discover new communication methods.

Fig. 3. (a) User using this application to recognize AR marker, and the latest Tweets will be presented. (b) The interface layout design. (c) The AR marker is printed on T-shirt on two sides.

3.6 Navigation

When users enter this application, they can use this iPhone application to recognize people who wear the T-shirt with AR marker. The latest Tweets will be loaded. Users can check the latest Tweets real-time on people's T-shirt, reply, and discuss it directly. They can only reply by using this application. If in a communication, people have some news have been found. They can also send the latest message via this system. And they can see the change happens on T-shirt in real-time, depends on the scenarios setting. We also have a discussion of using this application as following:

1. When users want to read AR marker, users can stand in front of people who wears the T-shirt with AR marker. After loading the latest tweet from twitter, they can start a new communication. In a communication, users also can reply message directly by using our system.
2. Because we also printed AR marker behind our T-shirt. Thus, users also can reply message for people wearing our T-shirt. This kind of interaction method in our system can also become a subject to study. If people got a message from strangers. These message directly presented on their T-shirt. This situation will also become a new experience when using social network service.

4 Conclusion

In this paper, we have proposed a new communicating method for users. User can use mobile phones to receive others' SNS information. In our current scenarios, the system is only designed for two users. For the future work, we propose an improved framework of the system that can be used by multiple users. Our future models will enable them to interact with more than two users. This study is the first step to our research in combining AR technology with wearable clothing. Using AR technology as the basis, we can set 2D/3D graphic, animation, video, or sound as communication contents to enhance this interacting system further.

References

1. Augmented Reality Technologies, Systems, and Applications. In: Multimedia Tools Application 2011, pp. 341–377 (2011)
2. Akioka, S., Kato, N.: Cross-media Impact on Twitter in Japan. In: SMUC 2010: Proceedings of the 2nd International Workshop on Search and Mining User-Generated Contents, Toronto Canada, pp. 111–118 (2010)
3. QR Stuff, http://www.qrstuff.com/
4. Panasonic (2010), http://panasonic.jp/viera/iphoneapp/promotion/Mmmmm
5. AR-Tees (2010), http://www.sebastianmerchel.de/artees.html
6. N-Building (2009), http://www.aircord.co.jp/blog/2010/01/n-building.php

Computer Support of Team Work on Mobile Devices

Hilko Donker[1] and Malte Blumberg[2]

[1] Dresden University of Technology, Faculty of Computer Science, Noethnitzer Str. 46, 01187 Dresden
[2] CoreMedia, Ludwig-Erhard-Str. 18, 20459 Hamburg
hilko.donker@acm.org, papers@maltebb.de

Abstract. In this paper we present a general concept of a mobile access to a groupware. The central aspect is how to bridge the gap between Mobile Computing and Collaborative Software. Mobile devices have limited capabilities, and therefore only few user interactions are desired. Conversely, groupware requires numerous interactions in order to make virtual collaborative work effective. First, we examine existing approaches and define our specific goal. Then, we present background on our research on user requirements. Afterwards, the general aspects of a prototype we developed are shown, including exemplary examples. After having given information about the first evaluation results, we end with a short conclusion stating our future work.

Keywords: Mobile Groupware, UI Design for mobile devices.

1 Introduction

Groupware systems are usually conceptualized as applications that enable computer supported cooperative work [4]. These systems support (small) groups of people who are involved in a common task and who work on common goals. Small group collaboration requires members coordinating their activities, recognizing the activities of other group members, and who are able to communicate with each other. Groupware systems have to provide a kind of "group feeling", called collaboration awareness [1] [3]. The participants have to be aware of other users involved in the collaborative task.

Today, users increasingly work in environments with varying resource constraints or where mobility is required. Mobile devices such as Personal Digital Assistants (PDAs), smart phones and cell phones have become a standard equipment of employees working at different locations or on the move. With the massive introduction of web-enabled mobile devices, users of this kind of devices are able to access a Groupware system from almost everywhere. Mobile devices are an appropriate medium for the delivery of important and just-in-time information. The use of such mobile devices leads to a new generation of Groupware systems, called mobile Groupware systems [11]. Extending stationary Groupware concepts to mobile devices offers great potential. However, too straightforward approaches, e.g. simply using desktop Groupware systems on mobile devices, fail due to the different nature of mobile devices and networks [14]. Obviously,

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 38–47, 2011.

a system cannot deliver the same (amount of) information to a mobile device that it delivers to a desktop device. It has to adapt to the users' context and deliver a reduced and adapted experience. Groupware systems significantly differ from single-user applications. Many users provide input (often simultaneously); output has to be processed for many users and shared data have to be kept consistent. This level of interaction is particularly challenging to support using mobile technologies when synchrony and timeliness of information is an issue [13]. In this paper our focus will not be on the support of loosely coupled group members, but on people with a high interdependence, so called teams.

1.1 User Interface Design of Applications on Mobile Devices

The market for cell phones, smart phones and Personal Digital Assistants (PDAs) is one of the most dynamic and competitive markets in the consumer devices industry. The input and output capabilities of cell phones are different from laptops or desktop PCs. In spite of rapid developments (e.g. iPhone), mobile devices still have more restrictions than the stationary equivalents. The term "handheld" implies that screen space is a limiting resource for interface development on such devices. Mobile interaction design has been recently addressed in a large number of publications (e.g. [7][9][17]). Zhang et al. [18] give an overview of the literature available on UI design, usability and related topics for mobile devices, in particular for cell phones and, to a smaller extent, for PDAs. They discuss that mobile devices have unique features which pose a number of significant challenges for examining usability, such as mobile context, connectivity, small screen size and restrictive data entry methods.

Considering the characteristics of mobile devices especially their pervasive and ubiquitous nature, the small size factor and the unique interaction modalities (e.g., touch screens, stylus, fingers, and combinations of the previous), a new range of usage paradigms has emerged [15]. As a consequence new usability guidelines have been established in order to provide users with applications that enhance their tasks and activities. Therefore, we put high effort on finding out the requirements for the mobile access to the Groupware.

1.2 Mobile Groupware

Roth et al. argue that although several Groupware systems are available, they can hardly be adapted to handheld devices [13]. Straightforward approaches, such as simply cross-compiling existing applications, fail owing to the specific properties of handheld devices and the connecting network: Handheld devices have low computational power, small memory and usually no mass storage devices. Handheld operating systems do not offer the same variety of services as desktop operating systems. Handheld applications follow a different usage paradigm. Network connections are still a problem. Litiu et al. point out that Groupware systems will dedicatedly need to be developed to work in such environments [12]. The variability occurs along several dimensions: user and application demands, user mobility and intermittent connectivity, hardware and network variability.

From our point of view mobile collaboration creates the possibility for users to be connected to their work environment while they are on the move. Some of the improvements are: informal meetings between meetings and working on spontaneous thoughts or urgent issues. Conclusively the attempt of not distinguishing between mobile scenarios and conventional situations is wrong. Mobile collaboration will not be possible in any given situation. It is necessary to evaluate the nature of mobile scenarios in order to decide, which situations are capable for mobile collaboration.

1.3 Work in Teams

Our objective is to realize a mobile computer support of the work in virtual teams. It is important to reflect on the different organizational roles in teams and on the user tasks in these teams. These roles are comparable in teams and virtual teams. It is essential to distinguish between different roles in collaborative software [16]. There are many different approaches when analyzing roles in teams [10]. We have limited our distinction to roles considering work positions and not tasks. The 3 roles we analyzed are: team member, team leader and customer or supervisor.

We believe that a team leader will derive much benefit from a mobile access to a Groupware. It is crucial for him to know the current state of the team's project so that he can react appropriately. He is involved in almost all essential decisions regarding the team's process. Therefore, he or she needs to be able to coordinate, communicate and collaborate promptly when urgent issues come up.

A mobile access will also create advantages for team members. The effectiveness of a team will increase, if team members have the opportunity to stay constantly in contact with each other. A mobile access will therefore help to create a strong togetherness in the team. Furthermore, team members can use a mobile access to react to urgent issues on short notice.

2 Requirements Analysis

We put high effort in finding out the requirements for the mobile access to the desktop Groupware JoinedForcedGroups. We based our assumptions for building the mobile application on 4 steps, finding out what needed to be supported and what did not. First, we needed to find out where and when a mobile access is needed. Working on mobile devices is only desired in certain situations. So, we had to do a substantial scenario analysis. Second the stakeholders needed to be identified, finding out who needed to be supported by a mobile access, who would benefit and who would not. Third after finding out who uses the mobile access, we had to examine the tasks of all identified stakeholders. Fourth and last we could conceptualize how the users' tasks needed to be supported. This step faced actual user actions, and therefore the actual development of the prototype could found on these results.

2.1 Scenarios of Mobile Work

A mobile access to software needs to fit useful mobile situations. The influencing factors describing such a mobile situation are:

- User Position (e.g. “in a train”, “walking”, “in a coffee bar”),
- Amount of distraction (e.g. “driving a car”, “attending a meeting”),

– Importance (e.g. “crucial project issue”, “customer request”),
– Available mobile device (e.g. “cell phone”, “laptop”),
– Connection, (e.g. “3G”, “WIFI”).

When a situation is examined, every factor has to be evaluated. As a result, the necessity for mobile work can be determined for every factor. A scenario is useful for mobile collaboration when no resulting necessity of a factor has been rated “not given”.

2.2 User Tasks

We needed to find out, what kind of tasks needed to be supported when conceptualizing the mobile collaboration access. We followed a three-step process: First we identified the general tasks coming up in a team’s lifecycle. Second we built a matrix combining these tasks with the functions a groupware offers to support collaboration. Third and last we examined which tasks were useful to support in a mobile scope. Our goal was to narrow down the entireness of team tasks to the most essential ones while the users are in a mobile situation. We wanted to be able to let the mobile access focus on user tasks with the most benefit. Of course, the nature of tasks in a mobile scope differs from the conventional ways of working with a Groupware. Thus, we wanted find out which tasks had the capability of being altered accordingly to make a mobile access to JoinedForcedGroups a benefit.

We restructured the user tasks to find out which ones we needed to focus on developing our mobile access. We chose the 3 categories “tasks that work both in a mobile environment and in a stationary environment”, ”tasks that had to be altered to work in a mobile environment” and ”tasks that do not work in a mobile environment at all”.

First we analyze the tasks of the team leaders:

– Tasks that could be directly adapted from stationary to mobile environment are e.g.:
  Spread information, develop ideas, rate ideas, quality assurance.
– Tasks that work differently in a mobile environment are e.g.:
  None.
– Tasks that do not work in a mobile environment e.g.:
  Compose teams, Give feedback on personal development of team members.

Second we analyze the tasks of the team members:

– Tasks that could be directly adapted from stationary to mobile environment are e.g.: Explain goals and assign tasks.
– Tasks that work differently in a mobile environment are e.g.:
  Communicate idea suggestions, read ideas, create and spread short pieces of information.
– Tasks that do not work in a mobile environment are e.g.:
  Present ideas, develop prototypes, build an infrastructure.

3 Concept and Prototype

There is one key factor when developing mobile Groupware access: Adapting the principles of collaborative work to mobile environments. Especially when an existing

Groupware system needs to be extended by a mobile access, several issues need to be faced.

The user experience of the stationary Groupware product needs to be transformed for mobile devices. Users who know e.g. JoinedForcedGroups should not have problems in switching between the desktop Groupware and the mobile equivalent. In order to ensure mobile usability, a drastic reduction of work steps needs to be done. Users should get their tasks done as fast as possible. As a consequence, there is a correlation between the importance of a collaborative task and the effort to execute this task in a mobile environment. Therefore, it is essential to focus on events like the central element of a mobile groupware access. Not minimizing but mobilizing is the way of adapting a stationary product for mobile work [2].

3.1 User Interface Guidelines

A mobile user interface has to be simple. In contrast to conventional computers (e.g. desktop PC, laptop) mobile devices still have elementary constraints regarding the user interaction. This range is from difficult text entries to small displays that can hold only little information [2]. Facing these restrictions we developed guidelines for our mobile user interface.

1. Focus on events: We built an event feed where users can easily see what has happened in their team since their last login. Users can be notified through different channels when a certain event occurs. Users have easy ways of reacting to an event. Each event holds a link that refers to the source where the user can interact appropriately.
2. Save space: The space on the top of each page is the most important one, since data shown there can be seen without scrolling. A user first wants to know the context before interacting. All information on collaborative work (e.g. new entries, message contents, events) is located on the top of the screen. Functions and notifications are always on the very bottom of the screen. In list screens we cut off long texts after some length, giving users the possibility to view the whole text by simply clicking on it.
3. Adapt Navigation to mobile needs: All areas are in hierarchical order, building a consistent network from higher to lower levels of screens. A user can always see what area he is using. We preserved a fixed space for direct hierarchical navigation.
4. Offer help functions: All screens hold help texts. Contextual help can be accessed by a unique key or short cut.
5. Offer administrational functions for mobile needs: Only essential functions for mobile work (e.g. notification settings, device settings) are available for administration. Login and logoff have been built as simple as possible. Users stay logged on if desired. Furthermore device and personal settings are stored automatically.
6. Keep the access consistent and unique: The stationary groupware's look-and-feel was adapted for the mobile access. Elements on the screen have their fixed places. Consistent symbols were used for fixed functions. Consistent hot keys were assigned.

3.2 Example Cases

The collaborative elements of our mobile access are much easier to follow when examples are given. Therefore, we picked out use cases of a team leader working on a moderation task and a team member collaborating by using the brainstorming module.

Fig. 1. Left: Mobile Adaption of the Module Manager, Right: Mobile Adaption of the Brainstorming Module

As team moderators need to fulfill further tasks, a special area for team management is available. In this case moderators set up modules for the team. They can define certain time spans for each available module. For instance, moderators can restrict the access to a brainstorming module only to team phases where general ideas have to be developed. Moreover, moderators can edit a virtual flip chart. This works as a central point of information. Anything written on the flip chart will be shown to all team members as soon as they log on to the team. Important updates or general information can be put here. The left screenshot in Fig. 1.shows the mobile representation of a so-called module manager. Moderators use this service to set up modules for the team. In this case, three modules have been activated for the according team, each one with different time spans. The moderator sees the overview, while having ways to edit, delete or edit modules. The underlying modular concept of JoinedForcedGroups offers a basic structure for collaborative work: For different tasks of the team different suitable modules are available. The Groupware offers quite a variety of modules, having its focus on the support of creative tasks. All actual collaborative work takes place in the according modules. When working on ideas, people use brainstorming as a creativity technique. The mobile adaptation of Joined-ForcedGroup's brainstorming module can be seen in the right screenshot in fig.1.

The module offers different sessions for different topics. In the mobile version, only the current session is shown on the screen. The user can scroll through the already given ideas or search for a specific idea by using the dedicated search box. Moreover new ideas can be submitted by using the according link. Again, the navigation area on top of the screen gives direct feedback on the current user's position.

4 Study

4.1 Method and Participants

In a study with 10 participants working in pairs on specific tasks we analyzed the usability and the acceptance of our mobile groupware front end. During the study each participant had to act as a team leader first. The team leaders had to prepare the work for their teams. Afterwards these preparations were taken into account for the work of the second person in the team in the role as a team member. The tasks of team leaders and the actual team work differed extensively so that several collaborative tasks were considered while working with the system. Each participant had to fulfill a coordination task, a communication task, a cooperation task, a moderation task an information search and configuration task.

After all participants' had acted both as a team leader and a team member all participants had to fill in a questionnaire. The evaluation of the usability was based on the criteria of ISO 9241 part 110. For the evaluation of the mobile front end of the Groupware we focused on the criteria "suitability for the task", "self-descriptiveness" and "conformity with user expectations". As a measurement instrument the IsoMetrics [5] questionnaires for these criteria were used. In order to measure the pragmatic and the hedonic quality of the prototype we used the Attrakdiff [8].

It was our objective to evaluate the mobile Groupware front end by participants coming from a suitable context. Therefore all participants had work experience in business jobs, they were experienced in team work, they worked in mobile scenarios and they had experience with Groupware. We had ten participants five female and five male. The age of the participants was between 27 and 35. Seven participants were employees and three of them were freelancers. Most of the participants used chats and messengers in their daily life. Other Groupware features were known and some of them were frequently used.

4.2 Findings

The realization of the "suitability for the task" of the mobile access was measured from "good" up to "very good" by the participant. 89% of the questions unitized "suitability for the task" were answered by a "good" and "very good". 3% of the questions were answered with an "average" and 3% with a "bad". Things which were rated into these areas are e.g. the positioning of interaction objects and information objects on the screen. The criteria "self-descriptiveness" was only rated with moderate ratings. Regarding the questions unitized "self-descriptiveness" 38% were rated „good“ and 31% „very good“. Although these results are quite good 15% of the questions were not answered. Problems were e.g. identified concerning questions whether items are disabled in some situations. Analyzing the three considered criteria of the ISO 9241, part 110 the criterion "conformity with user expectations" achieved the best results. Regarding the questions unitized "conformity with user expectations" 55% of the questions were marked with a "very good" and 36% of the questions with a "good". Concerning this criterion the unique design of the mobile access were mentioned as a positive item.

Fig. 2. Ratings of the pragmatic and the hedonic quality of the mobile access

Comparing the pragmatic and the hedonic quality the mobile JFG access we used AttrakDiff. The results are shown in Fig. 2.The mobile JFG was marked with a "desired". The small confidential square indicates the agreement of the participants concerning their impression.

The pragmatic quality was rated higher than the hedonic quality. We expected this result, as the main focus of our mobile access to the Groupware was supporting the users to fulfill their tasks. The results indicate that we achieved this objective. Nevertheless the ratings considering the hedonic quality of the software are absolutely positive. These results indicate that it was a joy for the users to work with the mobile access. A detailed analysis of the hedonic quality is shown in Fig. 3. Beside the pragmatic quality in this diagram the hedonic quality is divided into two sub-qualities, namely identity and stimulation. The sub-quality "identity" (HQ-I) indicates in which way a user identifies with the product. Within the sub-qualities the "identity" achieved the best results. The participants could identify themselves with mobile access even if they still recognized the pragmatic quality of the user interface.

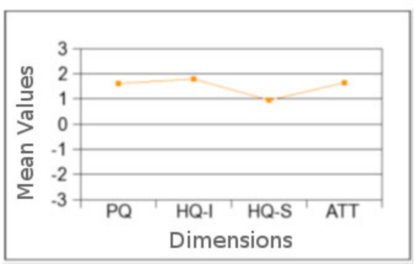


Fig. 3. Mean ratings of the dimensions "pragmatic quality" (PQ), "hedonic quality sub divided into "identity" (HQ-I) and "stimulation" (HQ-S), "attractiveness" (ATT)

The sub-quality "stimulation" (HQ-S) indicates the extent to which a product supports these needs by offering novel, interesting and stimulating features, contents, interactions and styles of presentation. The mean value of "stimulation" is less than the ratings of the other qualities. The users identified themselves with the user interface but they did not recognize it as novel or stimulating.

And finally the "attractiveness" (ATT) was measured. This is a subjective judgment of the attractiveness of a product. It is a matter of global rating on the basis of quality perception. Also the attractiveness was rated positive. In the detailed questionnaire the participants classified the user interface as presentable and invitational.

5 Conclusion and Future Work

The usability as well as the attractiveness of the mobile Groupware user interface were rated positive by the participants of our study. The user interface fits the needs of the users and it enables the users to interact with the Groupware working on their tasks while they are in a mobile situation. On the move it offers the users the additional benefit to be able to continue the collaboration with their colleagues. We realized an approach which enables people to establish an intensive virtual collaboration even if they were not able to meet physically or if they had no system with a large screen available. In a first step we focused on realizing a mobile user interface for some of the modules of a stationary groupware which are very promising to work on while on the move. These modules support typically an asynchronous collaboration. In a next step even synchronous modules have to be considered. Chat systems and voice over IP are state of the art on mobile devices. Concerning a mobile synchronous support of creativity techniques interesting research questions have to be analyzed. A ubiquitous mobile access to a groupware could be another promising approach for future work in this domain.

References

1. Beaudouin-Lafon, M., Karsenty, A.: Transparency and Awareness in a Real-Time Groupware System. In: Proceedings of the ACM Symposium on User Interface Software and Technology UIST 1992, Monterey, pp. 171–180. ACM Press, New York (1992)
2. Ballard, B.: Designing the Mobile User Experience. Wiley, West Sussex (2007)
3. Dourish, P., Bellotti, V.: Awareness and coordination in shared workspaces. In: Proceedings of the ACM Conference on Computer-Supported Cooperative Work, pp. 107–114. ACM Press, New York (1992)
4. Ellis, C.A., Gibbs, S.J., Rein, G.L.: Groupware: Some Issues and Experiences. Communications of the ACM 34(1), 38–58 (1991)
5. Gediga, G., Hamborg, K., Düntsch, I.: The IsoMetrics Usability Inventory: An operationalisation of ISO 9241-10. Behaviour and Information Technology 18, 151–164 (1999)
6. Geven, A., et al.: Always-on Information: Services and Applications on the Mobile Desktop. In: Proceedings of the 10th International Conference on Human Computer Interaction with Mobile Devices and Services, pp. 23–32. ACM Press, New York (2008)
7. Hagen, P., et al.: Emerging Research Methods for Understanding Mobile Technology Use. In: Proceedings of the 17th Australia Conference on Computer-Human Interaction, pp. 1–10. ACM Press, New York (2005)
8. Hassenzahl, M., Burmester, M., Koller, F.: AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In: Ziegler, J., Szwillus, G. (eds.) Mensch & Computer 2003, pp. 187–196 (2003)
9. Jones, M., Marsden, G.: Mobile Interaction Design. Wiley, West Sussex (2006)

10. Kauffeld, S.: Teamdiagnose. Verlag für Angewandte Psychologie, Göttingen (2001)
11. Kirsch-Pinheiro, M., Gensel, J., Martin, H.: Awareness on Mobile Groupware Systems. In: Karmouch, A., Korba, L., Madeira, E.R.M. (eds.) MATA 2004. LNCS, vol. 3284, pp. 78–87. Springer, Heidelberg (2004)
12. Litiu, R., Prakash, A.: Developing Adaptive Groupware Applications Using a Mobile Component Framework. In: Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work, pp. 107–116. ACM Press, New York (2000)
13. Roth, J., Unger, C.: Using Handheld Devices in Synchronous Collaborative Scenarios. In: Thomas, P., Gellersen, H.-W. (eds.) HUC 2000. LNCS, vol. 1927, pp. 187–252. Springer, Heidelberg (2000)
14. Roth, J.: Seven Challenges for Developers of Mobile Groupware. In: Workshop "Mobile Ad Hoc Collaboration" at CHI 2002 (2002)
15. de Sá, M., Carriço, L.: Lessons from Early Stages Design of Mobile Applications. In: Proceedings of the 10th International Conference on Human Computer Interaction with Mobile Devices and Services, pp. 127–136. ACM Press, New York (2008)
16. Schlichter, J., Borghoff, U.: Computer-Supported Cooperative Work: Introduction to Distributed Applications. Springer, Berlin (2000)
17. Weiss, S.: Handheld Usability. Wiley, West Sussex (2002)
18. Zhang, D., Adipat, B.: Challenges, methodologies, and issues in the usability testing of mobile applications. International Journal of Human-Computer Interaction 18(3), 293–308 (2005)

ProJest: Enabling Higher Levels of Collaboration Using Today's Mobile Devices

Babak Forutanpour and Jianfeng Ren

Qualcomm Incorporated
5775 Morehouse Drive, San Diego, California 92121
{babakf,renj}@qualcomm.com

Abstract. In this paper, we propose a real time and robust hand gesture system called ProJest which allows the combination of a smartphone or laptop paired with a pico or standard projector to turn any wall into a highly collaborative touch-capable surface. The system comprises of 4 stages: a calibration stage, followed by hand detection, hand tracking, and a gesture recognition step. Testing of the ProJest system show it working well under a variety of challenging environments, such as when users are interacting with content that has many skin-like colors or large amounts of motion.

Keywords: marker-less hand gesture recognition, collaborative workspaces, camera-projector space.

1 Introduction

Projectors are widely used today by individuals who need to share visual content with more than a few friends or associates. Projection systems must be connected to an input source, often a laptop computer, and require at least one individual to control the running application. This individual must often sit in front of the laptop to have access to the computer's mouse and keyboard in order to advance the presentation to the next or previous slide, highlight portions of the content, draw, zoom in and out of regions of interest, play or pause video clips, click on UI elements, etc.

Although being seated next to the computer is acceptable in many situations, there are times when the presenter would prefer to speak to his audience while facing all of them, standing next to the content being shared. Unfortunately, this would require the presenter to continuously walk between the screen and computer in order to control the running application.

It is therefore beneficial to develop a system which would allow the user to control the application while standing at the screen. It would be further advantageous to also allow audience members the opportunity to control the laptop from afar, in order to participate in the exploration of the content to drive higher levels of collaboration.

While there exist handheld remote controls that allow users to be un-tethered from the projector-computer rig, low to moderately priced versions have limited functionality and do not allow users to draw freely on the surface, highlight, copy, paste or interact with UI elements. Regardless of cost, such systems require every person in the

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 48–58, 2011.

room who wishes to be an active participant in the creation process to have their own device, something that may not always be possible.

Given humans often use hand gestures in everyday life as means of communicating with one another and moving and shaping objects in the world around them, allowing participants to user their hands in order to remotely control the application running on the PC would be the preferred method of interaction. See Fig 1.

Fig. 1. Depicts a user who may wish to become an active participant in the session by standing up and using his hand to draw, highlight, or zoom on the projected content. Here the content being projected is from a smartphone embedded with a pico projector.

In this paper, we describe a hand gesture recognition system for use in front of projection systems, called ProJest. Our system leverages the camera that has become common on many of today's laptops and mobile phones. While these cameras were included for teleconferencing or picture taking purposes, we use the onboard sensor as the system's input device in place of the mouse or keyboard.

In the coming years, we foresee even more opportunities for individuals to turn any wall into a highly collaborative workspace, given more of today's smartphones are not only equipped with cameras that could be used to track the hands, but pico projectors, capable of projecting 15 lumens, enough to create a 4 foot diagonal image in a moderately lit room. See Fig. 2.

Fig. 2. Qualcomm Snapdragon powered LG-Expo equipped with pico projector. The camera and projector on this and similar phones do not point the same direction, but this may change in the future.

In ProJest, the only conditions placed on the presenter are to ensure that: a) the laptop used to drive the projector is facing the screen such that the entire contents of the projected area are within the field of view of the onboard camera, b) a short calibration stage is completed before the session begins, and c) educating the audience on the two hand gestures required to drive the system.

We start by sharing with the reader several popular approaches in the field of hand gesture recognition. We then describe the four different stages of ProJest: a two-step calibration stage which maps coordinates from the projected images to computer's display space; followed by the determination of hand size; then a system which allows users to initiate the tracking process; followed by an algorithm developed to track the hand based on proposed new Haar features; finally, we describe a finger recognition system based on four Haar features which has been mapped to the left button on the computer's mouse. Live demos have shown that our system performs well for a variety of different content examples.

2 Related Work

Hand detection is an area of active research in computer vision, especially given its level of complexity. When compared to face detection [5, 12], where many good solutions exist today, building a hand detection and tracking system that works well in all environments has have been difficult for several reasons.

First, an open hand is roughly half the size of the human face with respect to a front facing camera, making feature tracking more difficult at the same resolution. At VGA resolution, tracking a hand from 10 feet across the room provides the algorithm with roughly 20 pixels in width for use by the recognition system. This is half the data available to face detection algorithms. What is more, unlike the face, hands have very few unique features, which are essential for good tracking. In the camera-projector scenario, in order for users at the screen to see the content they wish to interact with, and to be able to use their finger tip vs. finger nail to interact with it, they must stand with their backs to the camera. This requires ProJest to track the back of the hand, which unlike the face, has very few unique features. Additionally, the manner by which a human hand moves freely through space poses considerably more challenges than face tracking, given hands are attached to long limbs, whose multiple joints provide far greater degrees of freedom than the neck.

Many hand gesture systems have been proposed for vision-based interactions with computing devices. In [1], Mathias proposed a real time hand gesture system for mobility purposes. The main features used in [1] are KLT features, which are placed on "good-to-track" skin color spots from detected area. KLT features are named after Kanade, Lucas and Tomasi who found that a steep brightness gradient along at least two directions makes for a promising feature for tracking purpose. Our experiments with the author's algorithms as implemented in OpenCV's api functions, cvGoodFeaturesToTrack and cvCalcOpticalFlowPyrLK did not yield good results for situations where the user's hand was more than a few feet from the camera.

In [2], Nguyen proposed a more complex and computationally inefficient hand gestures recognition algorithm using pseudo two dimensional Hidden Markov Models. Such systems would not work well in the camera- projector scenario given our tests on the Hue changes of human skin under various projected content showed a wide gamut, see Figure 3.

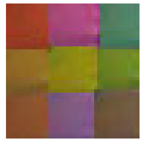


Fig. 3. 3x3 grid depicting skin under different projected light as seen by onboard camera. The wide color gamut makes skin-tone detection based methods ineffective.

The work appeared in [3] proposed a real time hand gesture algorithm for camera-projector system. During the calibration stage, the perspective distortion is modeled by polynomial warping between the coordinates of the camera and projected screen. Detection and tracking are based background segmentation. However, such a system requires static backgrounds, something that is not guaranteed in our system.

Given our motivation to design a system that can achieve real-time performance on a modern day embedded processor found in today's mobile phones, such as Qualcomm's Snapdragon core, we chose not to expand upon the work by Licsar and Sziranyi in [4] due the systems level of complexity. The requirement to read two VGA images, one being the projected image and the other how the camera sees this image on the screen, crop and warp and color correct the image from the camera for proper registration to the actual image, only to then search for multiple hands in the silhouettes created by the difference image did not seem feasible in under the 50ms required for perceived real-time performance.

In order to not burden the user, we did not consider systems which rely on time-of-flight cameras [6], other types of systems which require cameras working under infrared light [7], stereo cameras [8], or those that require the user to wear a special glove [9] or place colored caps on their fingers [10].

3 ProJest System

A key factor in getting users engaged in any new activity is to provide a low learning curve to some aspect of the system whereby the user can instantly experience some sense of pleasure. When the user's emotions come down from this initial feeling of joy, they may then decide to spend more time learning the system, perhaps even becoming a power user.

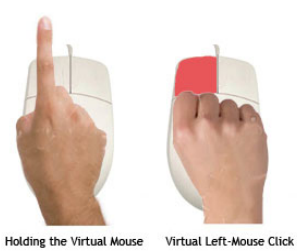


Fig. 4. The two supported gestures in the ProJest system

In order for our system to have minimal training for new users, yet provide high level of functionality, we chose to design the system around the way users already use their computers: when seated behind their desks. Here, users first reach for the computer mouse, place their hand over it, move it to some location on the table which maps the cursor to a desired location on the display, and then press one of the mouse buttons to accomplish some desired task. ProJest works similarly. Users must first walk up to the projected display, “pick up” a virtual mouse from some known location designated as a virtual “mouse pad”, move their hand over the screen to the desired location while the cursor follows their hand, and then “click” or “click and hold and drag” or “release click” on UI elements.

ProJest requires users to learn two intuitive gestures: a closed hand with extended index finger to simulate a user resting his hand on a computer mouse with his index finger on the left mouse button, and the same gesture but with the index finger hidden, simulating a click or clicked-and-hold operation. Releasing the mouse button is accomplished by revealing the finger in order to return to the resting position. Please see Figure 4.

In addition to being easy to learn, it is important that the system be real-time and accurate. Our goal was to have a system that was able to locate the hand position in under 50ms, and be accurate enough that a user can select UI elements on Microsoft Windows Vista’s Start Menu in a typical projection setup, where the projected image is 90 inches wide by 65 inches tall. Furthermore, in order to encourage users to host such collaborative sessions, the system must require zero or minimal calibration, and should not require the use of any equipment beyond a projector, laptop or phone with a VGA resolution camera. Our developed ProJest system includes a calibration stage, hand detection, hand tracking and gesture recognition stage as described below.

3.1 Calibration

Calibration is the first step for camera-projector pair systems. The coordinates of the projected content captured from the camera’s point of view, referred to as projected space, must be mapped to the those on the PC’s display, referred to as desktop space coordinates. See points A,B,C,D and A’, B’, C’, D’ in Figure 5 below. Images captured by the camera will not only suffer from location and scale differences relative to the desktop space coordinates, but also from perspective distortions because the objects are 2D projections of a 3D environment.

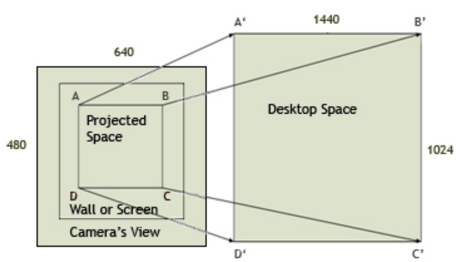


Fig. 5. Mapping coordinates from projector space to computer’s desktop space

In order to generate the transformation matrix, ProJest employs a one-time calibration stage. The calibration process displays a known image for projection, here a chessboard as seen in Figure 6. We use OpenCV’s cvFindChessboardCorners() to find the four corners of the chessboard, and then user these points to search for the actual coordinates of projected space. Naturally, for the calibration process to complete successfully, the content being projected must lay entirely within the camera’s field of view.

Knowing the hand size is another important step for accurate tracking. The predominant factor that determines the size of the hand as seen by the camera is how large an image the projector can create. The more powerful the projector, the further back it is placed from the screen to create the largest image possible, which accordingly requires the camera to be placed further away from the screen. As shown in Figure 7, we project an image with circles of different radii and ask the user to put his closed fist inside the circle that provides the tightest fit for 2 seconds. Based on changes in intensity, we identify which circle the user selected.

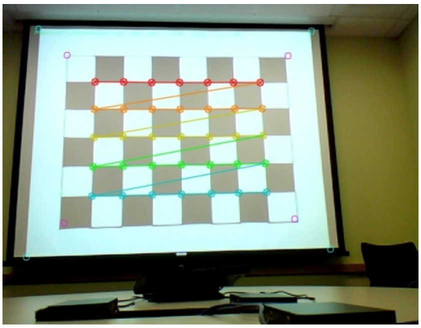


Fig. 6. The calibration process finds the corners of the projected space, denoted with Cyan circles, to the laptop's desktop

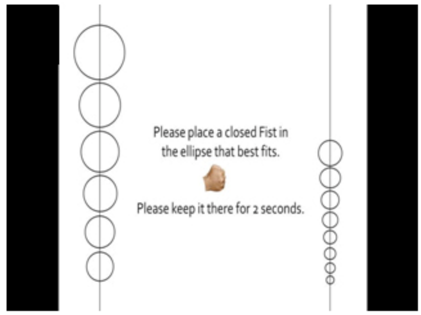


Fig. 7. Hand size determination calibration chart

3.2 Hand Detection

In our system, we ask the user to place his hand in a marked area of the projected space, referred to as the virtual "mouse pad." This step bootstraps the hand tracking process. Currently, the virtual mouse pad is in the bottom right corner of the screen and is 4 times larger than size of the hand that was determined during the calibration phase. While ProJest system is running, it is continuously differencing incoming pixel values in this marked area from corresponding pixels in previous frames. When the sum of differences as observed in the luminosity channel is above a fixed threshold, the system enters the hand tracking phase which starts by ensuring that the change is in fact due to a hand.

3.3 Hand Tracking

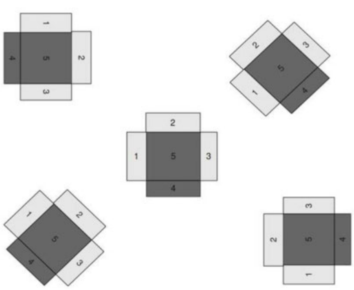


Fig. 8. Extended Haar features used to track the hand and wrist

Through analysis of captured video, we found that human skin absorbs more light than materials found in surrounding areas, such as screens, painted walls, white boards, etc. Given this, we propose 5 new Haar features depicted in Figure 8 which look for an object that has reflectance properties similar to the back of the hand and wrist under projected light relative to non skin materials.

We apply each of the five different filters on the data that covers an area 4 times larger than the measured hand size, at the last location in the image known to have been a hand. In this search space, we scan from left to right, top to bottom, pixel by pixel and select the pattern which has maximum difference between the sum of light gray areas for a given filter in Figure 8 and sum of corresponding dark gray areas. The filter with the best fit is then checked against a set of thresholds computed offline. More specifically, we check to see that the average pixel values in regions "1","2" and "3" are larger than those in area "5" plus one threshold, currently set to 20. At the same time, the average pixel values in area "4" must be less than the value in area "5" minus one threshold, again 20. Additionally, the average pixel value from center area "5" should be larger than 30, which helps us avoid tracking the head or shadows of the hand. If none of the 5 filters were successful in finding a fist and wrist, we repeat the process until a hand is found.

For real-time performance, integral images are computed for use by the 5 Haar features in Figure 8. For the three horizontal and vertical Haar features, integral images can be computed via the approach in [5]. For the two filters at +45 and -45 degrees, integral image are computed via rotated sum area table described in [11]. We terminate the search once one of the 5 filters matches the criteria.

3.4 Gesture Recognition

ProJest currently supports two different input actions: the press and release of the laptop's left mouse button. As depicted in Figure 4, a closed fist with extended index finger denotes when the mouse button is not depressed, while hiding the index figure maps to the button being pressed.

Once a tracked hand has stopped moving for some interval, e.g. 700ms, our system begins a search for the presence of the index finger. The system uses 4 Haar features for this, as shown in Figure 9. The gesture recognition begins by first scaling the tracked hand image into an image of fixed size, 40x40. Based on this image, integral image is computed for use when index finger is pointing in the horizontal or vertical direction. The criteria for the selected index finger is that average pixel value for "1" and "3" area in Figure 9 should be larger than the value in "2" area plus one threshold, e.g. 30. If a finger in the horizontal or vertical pattern is found, we terminate the search; otherwise, we compute the rotated integral image at +45 and -45 degrees. If no index finger is found using these to filters, we consider the hand in the closed fist gesture.

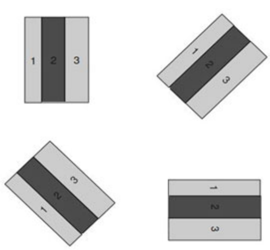


Fig. 9. Extended Haar features used for hand gesture recognition

In order to improve the robustness of the finger recognizer, we apply a maximum voting algorithm to a buffer of 20 stored frames. We require that at least 80% of the images in the buffer depict a hidden index finger before we trigger a left mouse button event to the OS's Window Manager. In order to detect a revealed index finger, 50% of the 20 frames searched must report the discovery of the index finger. The lower threshold requirement for the reveal vs. hiding of the index figure is due to our state machine which requires up mouse events to be triggered only after a down mouse event has been detected, which by itself serves as a filter to reduce false positives.

4 Results

The system was developed in C leveraging the OpenCV library. The laptop used was a HP Compaq 6710b configured with Intel Core 2 Duo T7300 2GHz processors, 2 GB RAM, running Windows XP. The webcam was a Creative Live! Cam Video IM Ultra capable of capturing VGA resolution images as 30fps. The projector was an Epson PowerLite 83+ projecting an image on a screen that measured 92 inches wide by 65 inches tall. The lighting in the room was lowered to 66 lux, where users could easily see the screen and one another comfortably. The light at the screen when projecting gray image having rgb value of 128 was 72 lux.

During the hand tracking and gesture recognition phases, the system was measured to consume only 23% of the CPU. To measure the accuracy of the system, we tabulated the number of times a user already familiar with the system could stand 10 feet away from the camera, and use his bare hand to navigate to 4 pre-determined positions on the screen in order to trigger a down mouse event followed by up mouse event.

The content being projected is a video playing in full-screen mode. To ensure the videos were chosen at random, we selected the 50 most popular videos on YouTube in the "Videos Being Watched Now" category at the time of the test. None of the videos were used during the system's development phase. A list of the URLs for the 50 videos is available upon request. To ensure the videos played throughout the test uninterrupted, we only selected videos that were longer than 1 minute in length, were fully buffered before the tests began, and disabled the gestures from actually triggering mouse events which would have paused the video.

The four regions on the screen are labeled A, B, C, D and map to the upper left, upper right, lower right, and lower left corner, respectively, of a rectangle that has ½ width and ½ height of the projected area centered at the bottom 1/3 of the screen,. The test required the user to pick up mouse from the virtual mouse pad, navigate the hand to position A, and if the cursor was successfully tracked to within 2 inches of that position, click on the region. If that gesture was recognized correctly, the user would then release the virtual mouse button, by revealing the index finger. Following this, the user would then move his hand to the right to position B, down to position C, and finally left to position D, doing the two hand gestures at each of these three positions as well. In all, 12 measurements were made for each video.

Table 1 shows the results of our 4 tests for each of 50 videos. We can see that ProJest was 96% effective in tracking the hand to each of the predetermined locations. Of the 192 times that the cursor tracked the hand successfully, it detected the click gesture 157 times, or 82% effective. Of these 157 successful recognitions, ProJest detected the accompanying release gesture 154 times.

Table 1. Ability for ProJest to track hand to 4 fixed position and recognize click and un-click gesture while one of 200 random videos from YouTube were selected

	Track	*Click*	*Release*
Attempts	200	192	157
Failed	8	35	3
Effectiveness	96%	82%	98%

The 82% effectiveness of the click gesture recognition reveals that there is room for improvement in our finger detection algorithm. However, in an effort to choose gestures that would be easy for users to both remember and learn, we exposed ourselves to having to track a finger that is 6 pixels wide given the VGA camera is 10 feet from the screen. In future versions, we may opt to map the click gesture to an action that is more visible from these distances, possibly an open to closed fist action. Optionally, we could leverage the robustness of the hand tracker to drive gesture engine. For example, after the hand has stopped for a period of time, the cursor is fixed to its current position, while the user is then able to move his hand left and right 3 times to denote a click action. Moving the hand right to left 3 times would denote a click and hold action, followed by right to left 3 more times for release of the button. In short, there is often an inverse relationship between the quality of gesture system and the intuitiveness of the gesture set. The simpler the gesture for the user, often the more complex it is to support.

The above results are in line with the feedback provided at Qualcomm's QCT Multimedia R&D Demo Day, where over a hundred volunteers either tried the system for themselves, or watched the authors use the system with high levels of success.

Figures 10, 11 and 12 are still images from users interacting with the touch capable wall. Actions depicted include drawing, zooming into a picture which is rich in skin like colors, or pausing and playing a youTube video, respectively.

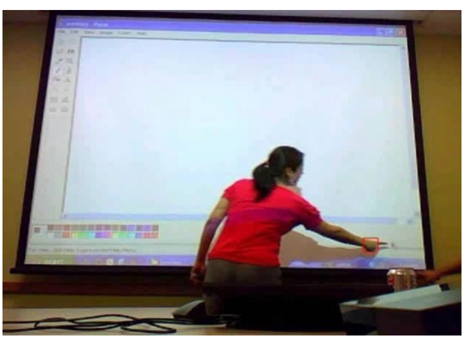


Fig. 10. User picking up the virtual mouse in order to draw pictures using MS Paint application

Fig. 11. User selecting an area of the image to zoom into. Note picture has abundant amounts of skin-like colors

Fig. 12. User "clicking" on screen to pause and play YouTube video

5 Conclusion

In this paper, we propose a real time and robust hand gesture system called ProJest which allows the combination of a Smartphone or laptop paired with a pico or standard projector to turn any wall into a highly collaborative touch-capable surface. The system comprises of 4 stages: a calibration stage, hand detection phase, followed by hand tracking, and finally a gesture recognition step. The system currently allows users to use their bare hands to control the single-button mouse of the laptop running the projected application from 10 feet across the room. After conducting 200 tests of the system, ProJest proved to be 96% effective in tracking the users hand; with 82% accuracy in detecting the gesture assigned to left mouse click, and 98% success rate of detecting the left mouse button being released. The system offers a real-time experience and only consumes 23% of the CPU on a modern day laptop. Despite the reasonable effectiveness of the hand tracking, we feel that a commercial system needs to work as well as today's input tools, e.g. computer mouse or remote control in order to be widely adopted. What is more, despite simplifying the calibration process, the authors feel that our system is still a bit too cumbersome given the user must setup the system in a position in the room where somebody may already be seated.

References

1. Turk, M., Kölsch, M., Höllerer, T.: Vision-based interfaces for mobility. In: 1st IEEE Intl. Conf. on Mobile and Ubiquitous Systems: Networking and Services, Boston, MA (2004)
2. Binh, N.D., Shuchi, E., Ejima, T.: Real-Time Hand Tracking and Gesture Recognition System. In: Proceedings of International Conference on Graphics, Vision and Image Processing (GVIP 2005), pp. 362–368 (December 2005)
3. von Hardenberg, C., Berard, F.: Bard-Hand Computer Interaction. In: Proceedings of the ACM Workshop on Perceptive User Interfaces (November 2001)
4. Licsár, A., Szirányi, T.: Hand Gesture Recognition in Camera-Projector System. Computer Vision in Human-Computer Interaction, 83–93 (2004)
5. Viola, P., Jones, M.: Rapdi object detection using a boosted cascade of simple features. In: IEEE Conference on Computer Vision and Pattern Recognition (2001)

6. Haker, M., Bohme, M., et al.: Deictic Gestures With a Time-of-Flight Camera. In: The 8th International Gesture Workshop (2009)
7. Wilson, A.: PlayAnywhere: a compact interactive tabletop projection-vision system. In: Proc. of the 18th ACM Symposium, User Interface Software and Technology
8. Malik, S., Laszlo, J.: Visual Touchpad: A Two-handed Gestural Input Device. In: Proceedings of the 6th International Conference on Multimodal Interfaces, pp. 289–296 (2004)
9. Wang, R., Popovic, J.: Real-Time Hand-Tracking with a Color Glove. ACM Transactions on Graphics 28(3), article 63 (August 2009)
10. Mistry, P., Maes, P.: SixthSense – A Wearable Gestural Interface. In: the Proceedings of SIGGRAPH Asia 2009, Sketch, Yokohama, Japan (2009)
11. Lienhart, R., Maydt, J.: An Extended Set of Haar-like Features for Rapid Object Detection. In: IEEE ICIP 2002, vol. 1, pp. 900–903 (2002)
12. Ren, J., Kehtarnavaz, N., Estevez, L.: Real-Time optimization of Viola-Jones face detection for mobile platforms. In: Proceedings of Seventh IEEE Dallas Circuits and Systems Workshop, pp. 1–4 (October 2009)

The Effect of Time Orientation and Representation of Points of Interests on the Use of Mobile Tour Guide

Fei Gao and Qin Gao

Tsinghua University, Beijing, 100084, China
gaof04@mails.tsinghua.edu.cn, gaoqin@tsinghua.edu.cn

Abstract. This study examined the effect of users' time orientation (polychronics and monochronics) and the display of points of interests (POIs) on users' workload and satisfaction when using a mobile tour guide. The results show that individuals' time orientation has a significant effect on users' workload. People who have a tendency of polychronic perceived higher workload. The display method of POIs significantly affected users' satisfaction. Categorized display was preferred by the users. For polychronics, display method had a significant effect, and altogether display was a better choice. Based on the findings, we suggest that mobile navigation interfaces design should pay extra attention to polychronic people as they tend to manage multiple activities simultaneously, which may add to mental workload. POIs in mobile navigation services should are displayed by categories, and an overall view with all POIs should be provided as well.

Keywords: Multitasking, time orientation, mobile.

1 Introduction

Mobile tour guide is becoming increasingly popular with new mobile devices embedded with GPS or other positioning system. Interaction with mobile devices is limited by its physical size, which results in restrictions on user interface and information presentation [4]. For mobile tour guide, a basic visualization task is to display points of interests (POIs) on the small screens of mobile devices. The selection of POIs being presented is critical because the limited space restricts how much information can be displayed on the screen.

Besides the limited size of screen, people's limited cognitive resources also make it difficult to interact with mobile devices. In outdoor environments, physical surroundings, other people, and mobile computing devices compete for cognitive resources [12]. Continuous attention to the mobile device was found fragmented and broke down to bursts of just 4 to 8 seconds, comparing to over 16 seconds in laboratory condition, and attention to the mobile device had to be interrupted by glancing the environment [12]. This implies that interaction with mobile devices is naturally multitasking. Previous studies [2, 7] showed that people are different from each other in their ability in handling multi-tasks and managing time, which was summarized by Hall [7] as time orientation. Time orientation, generally speaking, reflects the

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 59–68, 2011.

differences of people in managing time [7]. Some people tend to do one thing at a time (monochronic), while some others prefer to do several things simultaneously (polychronic). Will different styles of managing time influence people' usage of mobile applications? Shall special concerns be raised for designing interfaces for people with different time management style? When designing a mobile application, these issues should be concerned.

The aim of this study is first to examine whether the design of POI display will influence people's usage of mobile tour guide, especially when they have multiple destinations in their mind; second, whether there is any difference in satisfaction and workload between people with different time orientation when they use mobile tour guide. The interaction between time orientation and POI display is also investigated.

2 Related Work

2.1 Representation of POIs on Mobile Tour Guide

For mobile tour guides, one important visualization task is displaying a set of POIs. Usually, a specific icon is assigned to each category and drawn on a map in a way that visually maintains spatial relations among POIs. One major step to visualize information on mobile devices is selection, which refers to present the information relevant to the considered task [5]. Then how should we select the POIs to be presented in a mobile tour guide? On one hand, visualizing insufficient data will lead users to make suboptimal or plainly wrong decisions; on the other hand, burdening users with unnecessary data will make it more difficult to reason about a given problem. Although selection is an important aspect of any visualization, it is critical in mobile device visualizations because the limited space restricts how much information the screen can display. Chittaro [5] stated that the visualization should not simply draw all POIs as if they were equally relevant to the user. Instead, a typical approach is drawing only those POIs that satisfy users' needs. A frequently used method is display POIs by category.

2.2 Time Orientation

Time orientation describes the different ways people use in managing time. When facing abundant information, people's ability to manage time seems to influence their behavior. Activities are sometimes performed together (parallel) while at other times they are done one at a time (serially). This difference is summarized by Hall [7] as monochronic and polychronic time use. Monochronic people percept time as linear, they prefer to do one thing at a time; while polychronic people have a cyclic time perception and tend to do many things at once [7]. Monochronics tend to regard unscheduled events as interruptions [3].

Time orientation has an impact on people's behavior and preference. Bell et al.[1] proposed that polychronics are more likely to multitask with technology than monochronics. Ophir et al. [11] showed that heavy media multitaskers are more susceptible to interference from irrelevant environmental stimuli and from irrelevant representations in memory, thus, performed worse on a test of task-switching ability. The study of Huang et al. [8] revealed that polychronic users perceived lower level of interruption from the notification messages than monochronic users. Polychronic users preferred rapid and accurate response to the stimuli provided by the notification system while monochronic users tended to avoid that [8]. Lee and Harada [9] found that Japanese, with monochronic tendencies, preferred ‘deep’ interface structures and ‘verbal’ labels while Koreans and Americans, with polychronic tendencies, preferred ‘shallow’ and ‘graphic’ interfaces. This result may be attributed to information-overload as suggested by Haase et al. [6] who defined polychronicity as “the ability to cope with stimulus-intense, information-overload environments”.

Table 1. Comparison of monochronic and polychromic

Monochronic	Polychronic
Linear time perception	Cyclic time perception
Do one thing at a time	Do many things at once
Low tolerance for interruption	High tolerance for interruption
Feel “lost” or “disorganized” with excessive information	"live in a sea of information”

3 Research Questions

We proposed three research questions about the effect of time orientation and two display designs on users’ satisfaction and workload.

- *Question 1*: Will different display styles (categorized display and all together display) affect people’s perceived workload and satisfaction in using maps?

We compared two ways to display POIs on map: display by categories, which distributed different categories of information to different map layers and presented only one category at one time; the other way displayed POIs altogether. Due to the limited screen space, categorized display would be clear and organized. It will reduce users’ workload in visual search and increase the satisfaction. On the other hand, if different categories of information are separated, users have to dig “deeper” to get the information needed. It is difficult for the user to integrate different categories of information in the same geographical area. When the POIs are presented together, which is a “shallow” view, users can process many things at a less complex level. These affect users’ workload and satisfaction in using mobile tour guide.

- *Question 2*: Will people’s time orientation affect their perceived workload and satisfaction in using maps?

Time orientation affects users’ management of activities. Previous research on the characteristics of monochronics and polychronics implies that time orientation may influence people’s preference and performance when using mobile tour guide. Interaction while on the move is multi-tasking. People with different time orientation have different perception towards a multi-tasking environment.

- *Question3*: Will people with different time orientation have different preference for POIs display style?

The users have different requirements for information display. The amount of information displayed to a user at any one time may have an effect on performance depending on whether a person is monochronic or polychronic. Polychronics may want to process more information, as opposed to monochronics who may want only a limited amount of information related to only one task at any one time. Polychronics may found the all together display convenient to use. However, it will be distracting to monochronics to display all POIs together. They will consider the excessive information as useless and annoying.

4 Methodology

4.1 Experiment Apparatus

Two prototypes were developed for the experiment on ASUS A639 Pocket PC, which had 416 MHz CPU, 64MB RAM, 3.5 inch touch screen, 240×360 resolution and SiRFstarIII GPS module.

4.2 Participants

We conducted the experiment in the campus of Tsinghua University, which used to be an ancient royal garden and is now an important tourism site in Beijing. 40 participants who had never been to Tsinghua University before were recruited. 28 participants had bachelor's degree, 11 participants had master's degree and one participant's education level was high school. The ages of the participants ranged from 20 to 35 years old (mean=23.45, SD=2.679). 21 participants were male and 19 participants were female. The participants were young people with good computer skills.

4.3 Independent Variables

The two independent variables were POIs display style on map (display by categories or display all together) and time orientation score.

POIs display style. There were two POIs display styles: the categorized display of POIs and altogether display of POIs on map. Destination-related information was divided into four categories, while the basic layer of map showed basic geographical information like blocks and most important landmarks. The four categories were *sites of interests, education buildings, commercial facilities and traffic stops* respectively. Each category of information was denoted by a specific icon, as shown in Table 2. User's current position was marked by a red circle on the map. Participants operated the device using stylus. The map can be zoomed in, zoomed out and dragged to view different areas. User can use the "Return to current position" function to set the red mark back to the center of the screen.

Categorized display of POIs showed only one category of POIs at a time. For example, if user chose the category *sites of interests* from the menu, the interface would be as illustrated in Fig1(a), POIs of other type were invisible. When using the categorized display, user had to choose from a drop-down menu to select the category he/she wanted to view. Altogether display showed different categories of POIs all at once, as shown in Fig 1(b).

Table 2. POIs and number of points in each category

Category	Icon	No. of points
Sites of interests		7
Education buildings		5
Commercial facilities(restaurants and shops)		7
Traffic stops		1 routing, 2 stops

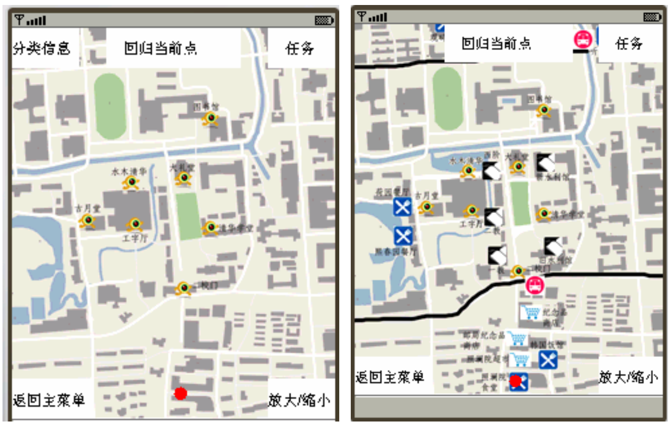


Fig. 1. (a) Categorized display; (b) Altogether display

Time orientation. We used the Inventory of Polychronic Values (IPV) in our experiment as an instrument for time orientation. Its reliability was 0.84[3]. According to IPV, the person is neutral (neither monochronic nor polychronic) if the calculated score is four. Score larger or smaller than four is cue for the tendency of time orientation. Individual IPV scores (M/P score) were used as continuous independent variable in analysis.

4.4 Dependent Variables

Data on workload and satisfaction were collected. We used NASA-TLX to measure users' workload. For the measure of satisfaction, we used End-User Computing Satisfaction (EUCS) [10] to assess users' satisfaction. Questions in this instrument are relevant and meaningful for measuring mobile applications. EUCS has five dimensions: content, accuracy, format, ease of use and timeliness. Reliability of the EUCS instrument was assessed in previous research by using Cronbach's α and found to be 0.928 for the test data and 0.938 for the retest data [10].

4.5 Procedure

We e-mailed the user the IPV questionnaire and calculated the score. 20 polychronic people and 20 monochromic people were recruited. Participants were randomly assigned to one display design until each design has 10 people with monochronic and 10 people with polychronic tendency. Then they were given a list of tasks and explained about the requirements of the tasks. The tasks were to find 10 specific destinations located within the campus. The tasks included: arriving at *(1) five sites of interests; (2) two education buildings; (3) two commercial facilities; (4) one traffic facility: the bus stop.* The users were given these tasks all at once to simulate a multitasking situation. These tasks did not need to be conducted by a pre-determined sequence. It was all up on the participants to determine their routes. Activities other than the given tasks were not allowed.

Before the experiment, the participants were introduced about the system and given about 5 to 10 minutes to get familiar with it. Then they were required to find a place not in the task list as training to ensure that they know how to operate the system. The user then began the tasks with the aid of PDA. The routing of the user and the time of the experiment were automatically traced and recorded by the prototype. At the completion of all tasks, the user filled questionnaires measuring workload (NASA-TLX) and satisfaction (EUCS).

5 Results

The impact of people's time orientation, the display method and their effect on people's workload and satisfaction was tested with structural equation modeling (SEM) method using Amos 16. The model was presented in Fig 2.

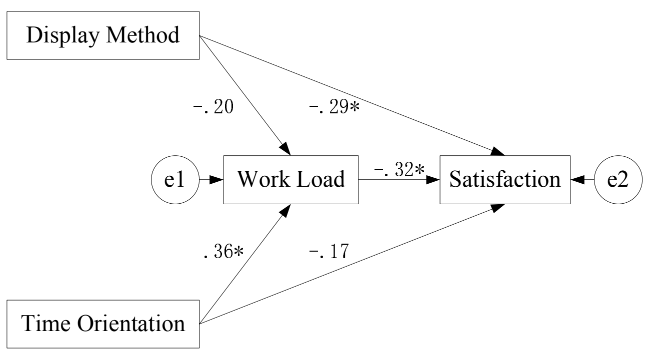


Fig. 2. Time Orientation, Display Method, Workload and Satisfaction

In this model, there were two exogenous variables: people's time orientation and the display method of POIs. Time orientation was measured using the IPV scale and used as a continuous variable. A higher score indicated a tendency for polychronic. The display method of POIs had two levels: 1 for categorized display and 2 for altogether display in the analysis. Endogenous variables were users' workload and satisfaction. Users' workload would affect their satisfaction. Workload was measured using NASA-TLA, with value from 0 to 100. Satisfaction was measured from a seven-point Likert scale (EUCS). A larger value indicated higher satisfaction. The structure equation of the above model was as below:

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ \beta_{21} & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} + \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} \zeta_1 \\ \zeta_2 \end{pmatrix} \quad (1)$$

Where y1 is workload measured using NASA-TLX; y2 is users' satisfaction; x1 is display method and x2 is people's time orientation.

Table 3. Standardized Regression Weights

			Estimate	p-value
NASA-TLX	←	Display Method	-0.200	0.171
NASA-TLX	←	Time Orientation	0.355	0.015
Satisfaction	←	Display Method	-0.286	0.049
Satisfaction	←	Time Orientation	-0.172	0.259
Satisfaction	←	NASA-TLX	-0.316	0.042

Chi-square value of the default model was 0.709, with degrees of freedom of one. The probability level was 0.400 > 0.05, which indicated the model presented above was not rejected. The fit indices were good. GFI (0.991) exceeded 0.9, CFI (1.000) reached the .95 standard, and RMSEA (0.000 < 0.05) was good.

Results for Question 1. Display styles of POIs significantly affect users' satisfaction (p=0.049), but its effect on workload is not significant. Categorized display (mean=4.08, SD=0.50) achieved higher satisfaction than altogether display (mean=3.90, SD=0.45). This may be resulted from the clear and organized presentation of categorized display. On small screen of mobile devices, presenting POIs altogether was a little bit crowded. We noticed that the difference on satisfaction resulting from two display styles was little. The reason for this may be that the total number of POIs in the experiment was not very large. Even if they were presented together, it still looks acceptable.

Although the effect on workload was not significant, there was a tendency that the altogether display would achieve lower workload, because the regression weight was negative. When the POIs were presented in category, changing from one category to another to find the next place to go is difficult. It required more effort to find information for different tasks using categorized display. If the POIs were presented all together, it is easier to find a routine with shorter distance.

Results for Question 2. Time orientation was found having a significant effect on NASA-TLX response. Polychronics (high M/P score) perceived higher workload, while monochronics (low M/P score) perceived lower workload (Regression weight = 5.27, Z = 2.43, p = 0.015). The reason for this may be that polychronics tended to manage different tasks simultaneously. Some participants asked if an overview can be

Table 4. Descriptive Statistics for Question 1

Style		**Mean**	**SD**
Categorized	Satisfaction	4.08	0.50
	NASA-TLX	44.85	12.28
Altogether	Satisfaction	3.90	0.45
	NASA-TLX	37.46	17.23

Table 5. Descriptive Statistics for Question 2 and 3

Time Orientation		Style	Mean	SD
Monochronic	Satisfaction	Categorized	4.21	.60
		All together	4.06	.37
		Total	4.13	.49
	NASA-TLX	Categorized	40.67	12.26
		All together	34.40	19.16
		Total	37.53	15.98
Polychronic	Satisfaction	Categorized	3.94	.35
		All together	3.74	.48
		Total	3.84	.42
	NASA-TLX	Categorized	49.03	11.38
		All together	40.51	15.47
		Total	44.77	13.92

provided when they were using categorized display. They tried to figure out a routine that can reduce their total walking distance, thus, increase the workload. Contrarily, Monochronics would rather make the next decision after finishing the current task. The effect of time orientation on satisfaction was not significant.

Results for Question 3. We tested the effect of display method within groups of participates with different time orientation by using MANOVA. Both NASA-TLX response and satisfaction were used as dependent variables.

The result showed that display method made a significant effect for polychronics (Wilks' Lambda=0.641, $p=0.023<0.05$). Since NASA-TLX response and satisfaction were correlated, the effect was mainly associated with one dependent variable. We assumed and proved that NASA-TLX response had a significant effect on satisfaction, thus, the difference was mainly resulted from workload. As presented in Table 7, for polychronics, workload of using categorized display (Mean=49.03, SD=11.38) was higher than using altogether display (Mean=40.51, SD=15.47). We considered that altogether display was a better choice for polychronics.

The effect of display method for monochronics was not significant. It cannot simply conclude which display method is better for monochronics. If reducing workload is more important, altogether display is better. While if achieving higher satisfaction is more important, categorized display is better.

There were some other observations during the experiment. We found that people's behaviors were affected by the application. When using such a handhold device, most people would keep on monitoring the screen of PDA to see his/her current position, and they may fail to notice the cues in the environment. This phenomenon was extremely observable when there was not much people and traffic on the road. Interaction with mobile device seems to influence people's situation awareness, which is defined as the perception of 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 [13]. This phenomenon requires further investigation, under different environment conditions, for its underlying reason and its implication on mobile application design.

Table 6. Test result for Question 3

TO	Statistic	Value	F	df	Error df	p
M	Wilks' Lambda	.933	.613	2	17	.554
P	Wilks' Lambda	.359	4.760	2	17	.023

6 Discussion and Conclusion

In this experiment, we simulated a multi-tasking environment by giving several tasks at once. By doing so, we examined the effect of time orientation and the display of POIs on users' satisfaction and workload.

Time orientation affected users' perceived workload. Since workload would negatively affect users' satisfaction, time orientation would affect users' satisfaction indirectly. People have a polychronic tendency perceived higher workload than those with monochronic tendency. This implies that polychronics does not necessarily more capable in handling multitasking environment. Following Bluedorn et al.[3] and Hall [7], polychronicity is the extent to which people (1) prefer to be engaged in two or more tasks or events simultaneously and are actually so engaged (the preference strongly implying the behavior and vice versa), and (2) believe their preference is the best way to do things. This implies that time orientation is more of preference than capability. In the study of Zhang et al. (2005), a positive correlation was shown between effort (one dimension of NASA-TLX workload measure) and M/P score. In other words, polychronics attempt to put in more effort possibly, in order to take care of the many different tasks at hand. The same reason may lead to the higher perceived workload of polychronics in our experiment.

In general, people felt higher satisfaction when using categorized display. Categorized display was preferred for its clearness of presenting on small screen. Admittedly, our finding about the preference of categorized display is valid assuming a limited number of POIs. If the number gets larger, altogether display may lead to visual crowdedness and consequently more visual search effort. The preference for categorized display would get stronger, or at least stay the same, comparing to more cluttered display. However, when users use the categorized display, they felt extra workload was brought in due to their unable to plan and organize the tasks efficiently as they expressed after the experiment. These call for a balance between workload and satisfaction when presenting information on mobile devices. We can conclude that display of the POIs on mobile tour guide should present information in category, and at the same time take the increased workload into concern. An overview of all POIs is helpful.

Polychronics tend to span their attention onto different tasks in a multi-tasking context. Their effort to take care of different tasks may increase the perceived workload. We need to pay more attention to them as most mobile computing creates multitasking environments, and they may be more vulnerable to such problems. The effect of time orientation on mobile computing application needs to be investigated in other scenarios. To suit the characteristics of polychronics and reduce the potential workload, tools can be designed to help polychronic users focus on a main task, providing notification from other tasks regarding to schedule, urgency and context.

References

1. Bell, C.S., Compeau, D.R., Olivera, F.: Understanding the Social Implications of Technological Multitasking: A Conceptual Model. In: Proc of the Fourth Annual Workshop on HCI Research in MIS (2005)
2. Bluedorn, A.C., Kalliath, T.J., Strube, M.J., Martin, G.D.: Polychronicity and the Inventory of Polychronic Values (IPV): The development of an instrument to measure a fundamental dimension of organizational culture. Journal of Managerial Psychology 14(3), 205–231 (1999)
3. Bluedorn, A.C., Kaufman, C.F., Lane, P.M.: How many things do you like to do at once? An introduction to monochronic and polychronic time. Academy of Management Executive 6(4), 17–26 (1992)
4. Chalmers, D., Sloman, M.: Survey of Quality of Service in Mobile Computing Environments. IEEE Communications Surveys 2, 2–10 (1999)
5. Chittaro, L.: Visualizing Information on Mobile Devices. Computer 39(33), 40–45 (2006)
6. Haase, R.F., Lee, D.Y., Banks, D.L.: Cognitive correlates of polychronicity. Perceptual and Motor Skills 49, 271–282 (1979)
7. Hall, E.T.: Understanding cultural differences. Intercultural Press, Yarmouth (1990)
8. Huang, D.-L., Rau, P.-L., Su, H., Tu, N., Zhao, C.: Effects of Time Orientation on Design of Notification Systems. In: Proc. HCI International 2007, Interaction Design and Usability, pp. 835–843 (2007)
9. Lee, K.P., Harada, A.: A study on the cultural effects on user-interface design: with the emphasis on the cross-cultural usability testing through world wide web. In: Proc. of the 4th Asian Design Conference (1999)
10. McHaneya, R., Hightowerb, R., White, D.: EUCS test ± retest reliability in representational model decision support systems. Information & Management 36(2), 109–119 (1999)
11. Ophir, E., Nass, C., Wagner, A.D.: Cognitive control in media multitaskers. In: PNAS (2009)
12. Oulasvirta, A., Tamminen, S., Roto, V., Kuorelahti, J.: Interaction in 4-second bursts: the fragmented nature of attentional resources in mobile HCI. In: Proc. CHI 2005, pp. 919–928. ACM Press, New York (2005)
13. Zhang, Y., Goonetilleke, R.S., Thomas Plocher, T., Liang, S.-F.M.: Time-related behaviour in multitasking situations. International Journal of Human-Computer Studies 62, 425–455 (2005)

The Virtual Workplace of a Mobile Employee – How Does Vischer's Model Function in Identifying Physical, Functional and Psychosocial Fit?

Ursula Hyrkkänen[1,*] and Suvi Nenonen[2]

[1] Aalto University and TUAS, P.O.Box 13300, 00076 Aalto, Finland
[2] Aalto University, P.O.Box 13300, 00076 Aalto, Finland
ursula.hyrkkanen@turkuamk.fi

Abstract. The article examines the applicability of Vischer's model of comfort and fit for classifying the features of virtual workplaces used in mobile work. The user-centered model of comfort and fit was applied in the context of systematic literature review. The review showed that the model of environmental fit is useful for more detailed classification of virtual places and spaces. However, it seems that in virtual work the threshold of workplace usability rises from the physical level to the functional level due to accessibility demands. A mobile employee is forced to completely stop working if he/she is not able to connect. Compared to Vischer's model the necessity level of the virtual workplace ascends to cover also the demands of functional fit.

Keywords: Virtual workplace, mobile work, comfort factors of virtual workplace.

1 Introduction

New developments in information and communication technology have changed the way people approach their life and work. Mobile virtual work is no longer bound to fixed locations as utilizing information and communication technology allows people to function freely in various environments. The employee is considered as mobile, when he works ten hours per week outside of the primary workplace and uses information and communication technologies for collaboration (1, 2). Virtual reality (3), as an environment relating to this new 'anytime anywhere work', can be called *the virtual workplace*. Virtual workplace provides connectivity through different size of devices and is accessed by different interfaces when supporting the performance of both individual and collaborative activities. Internet and intranet provide a platform for both simple (e.g. e-mail) and complex (e.g. collaborative working environments) communication tools. (4)

Our interest is the interrelationship between the physical and the virtual workplace not only in regard to their infrastructure, but also to their social and cultural contexts. Both the prerequisites connected to the virtual workplace and its actual use can be challenging. It could be claimed, for instance, that simultaneous physical and virtual

* Corresponding author.

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 69–75, 2011.

co-presence is generally not yet mastered in an effective way and that there still exist certain bottlenecks for a mobile employee entering virtual reality.

Vischer (5, 6) has analyzed the physical workplace as a physical, functional and psychological entity in order to identify features related to comfort and fit between a workplace and an employee (fig 1). When environment sets inappropriate or excessive demands to users, in spite of their adaptation and adjustment behaviours, it manifests the concept of misfit. In good fit there is a balance between a person's abilities, skills, degree of control and decision latitude and the work environment's demands, complexity, expectations and challenges. The nature of person-environment transaction arouses the sensation of either comfort or stress. Comfort may be considered as the fit of the user to the environment in the context of work (5, 7, see also 8).

According to Vischer (5), environmental comfort encompasses three hierarchical categories: physical, functional, and psychological. Physical comfort relates to basic human needs, i.e. safety, hygiene and accessibility. These needs are responded by applying building codes and standards. Functional comfort is defined in terms of support for users` performance of work related tasks and activities. Psychological comfort is related to feelings of belonging, ownership and control over workspace. We have expanded the category of psychological comfort and fit to cover also the social factors, and named the third category psychosocial comfort and fit.

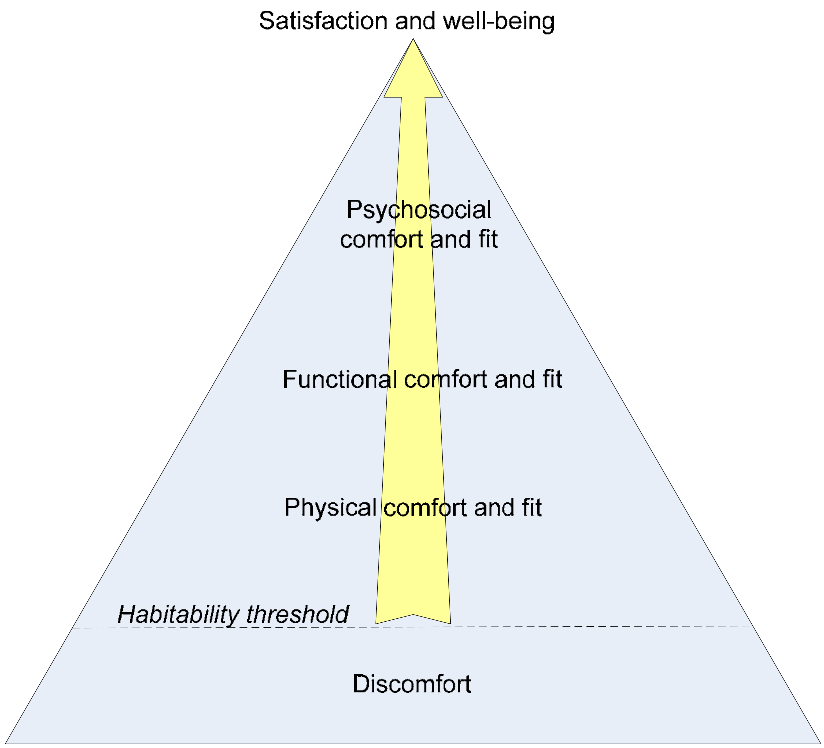


Fig. 1. Vischer's (7) habitability model of comfort and fit (modified)

Vischer's user-centered model merges environmental aspects with psychological aspects in a dynamic way. Vischer has developed a model for assessing the fit or misfit of physical workspace. The interesting question is whether the virtual workplace can also be captured as a three-level entity and whether this approach provides deeper understanding for managing it.

The purpose of this article is to explore how Vischer´s user-centered model of comfort and fit can contribute to identifying the different elements of the virtual workplace, which either hinder or enable productive mobile virtual work processes. By applying Vischer´s model, the research questions are:

1. What kind of physical comfort and fit elements can be identified in the virtual workplace of mobile employees?
2. What kind of functional comfort and fit elements can be identified in the virtual workplace of mobile employees?
3. What kind of psychosocial comfort and fit elements can be identified in the virtual workplace of mobile employees?

2 Method

The user-centered model of comfort and fit was applied in the context of systematic literature review. First, using a broad list of relevant terms, *virtual work, mobile work, nomadic work*, *virtual workplace* and *hybrid workspace* among others, a systematical search of several electronic multi-disciplinary databases (Scopus, Abi/Inform, Academic Search Elite, Elsevier Science Direct, Web of Science, Google Scholar) was conducted. Of the studies identified, only recent empirical articles (2001-2010) in peer-reviewed journals were selected and included.

Secondly, the outcomes of the 21 included articles were coded and classified according to Vischer's model. First all expressions concerning fit and misfit of mobile virtual work were identified. These expressions were then further classified to the levels of physical, functional and psychosocial factors.

3 Results

3.1 The Elements of Physical Comfort and Fit of the Virtual Workplace of a Mobile Employee

When considering the physical fit of virtual reality, fourteen of the reviewed articles described the constraints of physical places that mobile workers encounter. Constructions of physical environment blocked and hampered the employees' way to the virtual reality. Many odd places were offered for building up a work station (9), there were no large enough flat surfaces for devices (10), the physical infrastructure for mobile workers was neglected (11) and they had to compete for posts with local workers (11). There may have also been competition for electrical power, if there were not enough power outlets in public places (11, 12, 13, 14). The main reason for these misfit factors was the mobile and multi-local working mode. When executing the anywhere working style, the employee will undoubtedly encounter physical places that have not been are not in the first hand designed primarily for working purposes.

This is likely to happen for example at airports, in the different means of transportation, in cafeterias or in hotel rooms (12, 13, 15, 16, 17).

The physical fit of virtual reality is also a question of its appropriateness to the human sensory system. Especially visual and auditory problems were described in the reviewed articles. For ensuring the success of work, mobile employees carry many tools with them – they carry also redundant tools to be on the safe side. To avoid letting the weight of the burden grow beyond measure, increasingly smaller-sized devices are selected. With small size you inevitably also come to choose small displays - and visual difficulties. (2, 9, 12, 13, 18, 11, 19, 20, 21.)

Noisy environments may disturb concentrated working in virtual reality. Especially in public places, i.e. in trains and in airplanes, the tourists and neighbors beside the mobile worker may disturb the work (12, 15). On the other hand, a smooth level of discussing voices e.g. in a cafeteria may help the worker to relax and lose himself in virtual reality (4, 22).

Contradictory relation between physical and virtual worlds might cause misfit which may lead to safety risks, e.g. when driving a car (17, 20).

According to Vischer (5, 6), physical comfort creates the threshold for acceptable workplace. If a building does not meet the basic levels of demands stated in different environment and building standards and codes, it is uninhabitable and unusable. In comparison, the physical misfit of virtual place is caused by the contradictory relation between physical and virtual places and spaces. The main reason for this contradiction is the mobile working mode – i.e. the need to work anywhere. So far there are no standards, codes or rules that would specify the threshold for an acceptable workplace in a bus, train, hotel room or at home. There seldom are rules which would regulate the behaviour of people who use these places for work.

3.2 The Elements of Functional Comfort and Fit of the Virtual Workplace of a Mobile Employee

Relating to the functional fit of virtual places, fourteen of the articles presented connectivity problems that caused disturbances and hindrances to virtual work flow and broke it down. Some of the connectivity problems derived from the limited skills of mobile workers in employing virtual settings and infrastructure (2, 10, 11, 21, 23, 24). Time constraints and tight schedules of mobile employees together with time taking downloads of connections and programs also made it unreasonable to start virtual work (11, 12, 13, 15, 19, 20). The security regulation of mobile employees` own or their customers` company was reported to hinder the access and functioning in the virtual places (11, 13). In some cases the connections were also very expensive and therefore the access was forbidden (12).

According to Vischer (5), the functional fit or misfit of the workplace can be assessed by defining the degree to which occupants can either conserve their attention and energy for their tasks or expend it to cope with poor environmental conditions. In addition, the review observations showed that the functional misfit of a virtual place may lead to total breakdown of work flow. So the threshold of acceptable virtual workplace seems to lie not only at the physical level but also at the functional level.

3.3 The Elements of Psycho-Social Comfort and Fit of the Virtual Workplace of a Mobile Employee

Fifteen articles highlighted problems in the psycho-social fit of virtual place. The problems could be classified into six broad themes, which included controlling simultaneous co- and telepresence (11, 13, 23, 25, 26), belonging to physical and virtual communities (13, 23, 24), need for privacy (9, 10, 11, 12, 13, 15), expectation of ever-availability (11, 13, 18, 23, 25, 27) and problems of spreading tacit knowledge (23).

In Vischer's environmental comfort model, psychological comfort links psychosocial aspects with the environmental design and management of workspace through the concepts of territoriality, privacy and control (5, 7). A sense of territory is associated with feelings of belonging and ownership. Privacy is best understood as the need to exercise control over one's accessibility to others. Environmental control consists of mechanical or instrumental control, and empowerment (7). Instrumental control exists, if the employee masters his furniture, devices and tools. Empowerment as a form of environmental control arises from participation in workplace decision making. When comparing the factors identified from the reviewed articles to Vischer's psychosocial factors, the similarities are evident. Ensuring the psychosocial fit of a virtual workplace is the question of territoriality, privacy and control.

4 Conclusions

This review showed that Vischer's model (5, 7) of environmental fit is useful for more detailed classification of virtual places and spaces. However, it seems that in virtual work the threshold of habitability rises from the physical level to the functional level due to accessibility demands. The work of a mobile employee will totally stop if the worker is not able to connect. At the same time it should be realized that the concept of usability describes better the threshold of work in virtual world than habitability. So it seems that compared to Vischer ´s model the necessity level ascends to cover also the demands of functional fit (see fig. 2).

The use of the model also made it evident that in order to develop well-functioning virtual workplaces for mobile employees, broad attention should be paid to the whole system, within which employees confront their duties on different locations. Gaining comprehensive understanding about the context in which a given task is performed starts by forming questions first on physical place and later on psychosocial themes. Being such a vast field, the process demands profound multidisciplinary collaboration of different actors in organizations and support functions. The inspection of different fit levels is a useful tool for helping different authorities to explain their expertise in relation to other authorities. Gaps in management may also be demonstrated.

According to this review with Vischer ´s model as a frame of reference, it can be stated that

- at the level of physical fit, building codes and standards should be expanded to cover also the needs generated from the new working modes i.e. mobile work
- at the level of functional fit, the access creates the threshold of work. Moreover, the questions concerning the infrastructure, easiness of connecting signals as well as of finding help and support in using information technology are essential

– enhancing the fit at the psychosocial level: the mixture of physical and virtual worlds and simultaneous existence in both should be more effectively understood and supported as well as the integrated design, which seamlessly combine the physical and the virtual places need to be developed further.

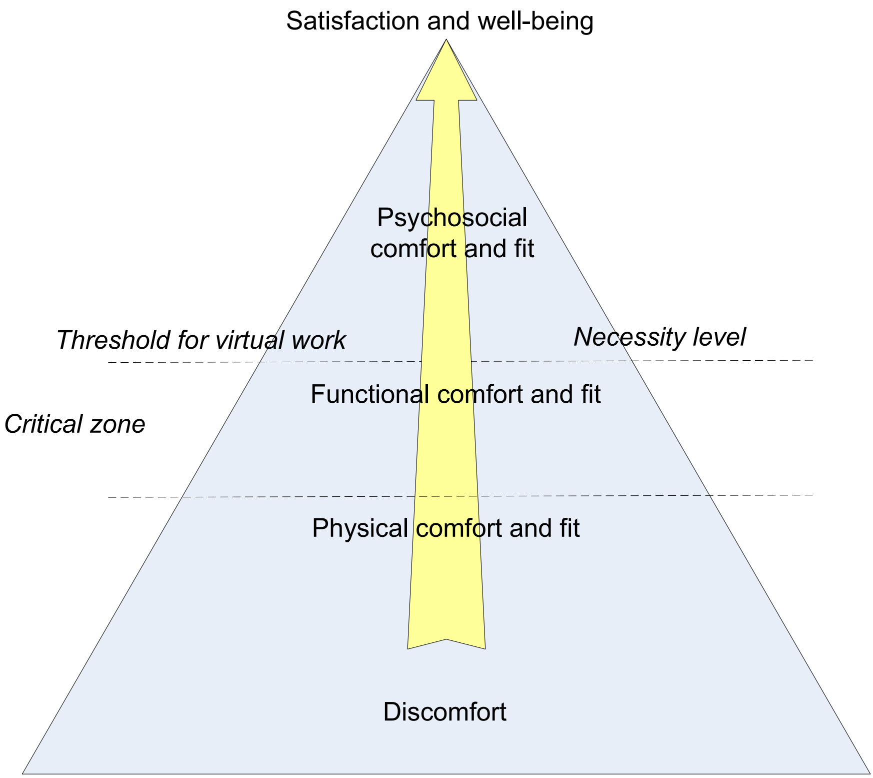


Fig. 2. Vischer's (2005) model of comfort and fit modified for assessing virtual work places

References

1. Gareis, K., Lilischkis, S., Mentrup, A.: Mapping the Mobile eWorkforce in Europe. In: Andriessen, J.H.E., Vartiainen, M. (eds.) Mobile Virtual Work. A New Paradigm?, pp. 45–69. Springer, Heidelberg (2006)
2. Vartiainen, M., Hyrkkänen, U.: Changing requirements and mental workload factors in mobile multi-location work. New Technology, Work and Employment 25(2), 117–135 (2010)
3. Fox, J., Arena, D., Bailenson, J.N.: Virtual reality. A Survival Guide for the Social Scientist. Journal of Media Psychology 21(3), 95–113 (2009)
4. Nenonen, S., Airo, K., Bosch, P., Fruchter, R., Koivisto, S., Gerberg, N., Rothe, P., Ruohomäki, V., Vartiainen, M.: Managing Workplace Resources for Knowledge Work, the final report of ProWork-project (2009), `http://www.proworkproject.com/prowork/final-report`
5. Vischer, J.C.: The effects of physical environment on job performance: towards a theoretical model of workspace stress. Stress and Health 23, 175–184 (2007)

6. Vischer, J.C.: Towards an Environmental Psychology of Workspace: How People are Affected by Environments for Work. Architectural Science Review 51(2), 97–108 (2008)
7. Vischer, J.C.: Space meets status. Designing workplace performance. Routledge, New York (2005)
8. Dainoff, M.J., Mark, L.S., Ye, L., Petrovic, M.: Forget About Aesthetics in Chair Design: Ergonomics Should Provide the Basis for Comfort. In: Dainoff, M.J. (ed.) HCII 2007 and EHAWC 2007. LNCS, vol. 4566, pp. 19–25. Springer, Heidelberg (2007)
9. Hislop, D., Axtell, C.: To infinity and beyond?: workspace and the multi-location worker. New Technology, Work and Employment 24(1), 60–75 (2009)
10. Perry, M., Brodie, J.: Virtually connected, practically mobile. In: Andriessen, Vartiainen (eds.) Mobile Virtual Work, pp. 97–127. Springer, Heidelberg (2006)
11. Mark, G., Su, N.M.: Making infrastructure visible for nomadic work. Pervasive and Mobile Computing (2010), `http://www.elsevier.com/locate/pmc`, doi:10.1016/j.pmcj.2009.12.004
12. Axtell, C., Hislop, D., Whittaker, S.: Mobile technologies on mobile spaces. Findings from the context of train travel. International Journal of Human-Computer Studies 66, 902–915 (2008)
13. Brown, B., O'Hara, K.: Place as a practical concern of mobile workers. Environment and Planning 35, 1565–1587 (2003)
14. Forlano, L.: Working on the move. The social and digital ecologies of mobile work place. In: Hislop (ed.) Mobility and Technology in the Workplace, pp. 28–42. Routledge, New York (2008)
15. Breure, A., van Meel, J.: Airport offices: Facilitating nomadic worker. Facilities 21(7/8), 175–179 (2003)
16. Laurier, E.: Doing office work on the motorway. Theory, Culture & Society 21(4/5), 261–277 (2004)
17. Laurier, E., Philo, C.: The region in the boot: mobilizing lone subjects and multiple object. Environment and Planning D 21, 85–106 (2003)
18. Felstead, A., Jewson, N., Walters, S.: Changing Places of Work. Palgrave Macmilan, New York (2005)
19. Perry, M., O'Hara, K., Sellen, A., Brown, B., Harper, R.: Dealing with Mobility: Understanding access anytime, anywhere. ACM Transactions on Computer-Human Interaction (TOCHI) 8(4), 323–347 (2001)
20. Perry, M., Brodie, J.: Virtually connected, practically mobile. In: Andriessen, Vartiainen (eds.) Mobile Virtual Work, pp. 97–127. Springer, Heidelberg (2006)
21. Venezia, C., Allee, V.: Supporting mobile worker networks: components for effective workplaces. Journal of Corporate Real Estate 9(3), 168–182 (2007)
22. Forlano, L.: Anytime? Aywhere?: Reframing Debates Around Municipal Wireless Networking. The Journal of Community Informatics 4(1) (2008), `http://ci-journal.net/index.php/ciej/article/view/438/401`
23. Hallford, S.: Hybrid workspace: re-spatialisations of work, organization and management. New Technology, Work and Employment 20(1), 19–33 (2005)
24. Mann, S., Holdsworth, L.: The psychological impact of teleworking: stress, emotions and health. New Tecnology, Work and Employment 18(3), 196–211 (2003)
25. Green, N.: On the Move. Technology, Mobility, and the Mediation of Social Time and Space. The Information Society 18, 281–292 (2002)
26. Tietze, S., Musson, G.: Recasting the home-work relationship: A case of mutual adjustment? Organization Studies 26(9), 1331–1352 (2005)
27. Tietze, S., Musson, G.: When work meets home. Temporal flexibility as lived experience. Time & Society 11(2/3), 315–334 (2002)

CornerPen: Smart Phone Is the Pen

Bong-gyu Jang, Myonghee Lee, and Gerard J. Kim

Digital Experience Laboratory
Korea University, Seoul, Korea
gjkim@korea.ac.kr

Abstract. The use of finger on the touchscreen is one of the most prevalent forms of input on mobile devices. However, due to the size of the finger tip, precise input is difficult and the presence of the finger on the touchscreen can often occlude the content in interaction. In this paper, we propose to install a sensor in the corner of the mobile device (e.g. smart phone) and transform the mobile device into a digital pen for making input for itself or other external devices. The system, dubbed "CornerPen" has many potential advantages in addition to those of the traditional pen-based input (vs. finger based) such as less occlusion, leveraging on tactile memory, and larger interaction surface. We have implemented and experimentally tested the CornerPen against the nominal finger-based touchscreen input system using two tasks, namely, flick-based icon browsing (search) and selection and free-form text input. Our results showed while the subjects did acknowledge the problem of occlusion with finger-based input on the touchscreen, the CornerPen approach still was not particularly effective nor preferred for the intended purpose, i.e. making precise input, and only exhibited comparable performance for simple flick/tab like input actions.

Keywords: Mobile interaction, Optical tracking, Tactile memory, Pen-based interaction, Finger touch, Icon selection, Usability.

1 Introduction

Making input using a pen-like device (e.g. stylus, digital pen) is one of the most natural and precise interaction methods. As such, many realizations of pen-like and "direct"[1] input methods have been developed such as the light pen, digital pen, stylus-touchscreen, and finger-touchscreen[2]. In particular, many mobile devices use the touchscreen and its finger/pen based input generally suffers from the occlusion problem (the finger/hand occluding the display). The finger touch input also often fails to provide relatively high precision input [1][2].

[1] The mouse can be thought of as a generalization of a pen, but implemented as an indirect interaction device. Digital tablet uses a pen that provides indirect control of a cursor. Note that a pen-like device can carry out virtually all the input possible with the mouse, but usually with higher precision (due to the nature of the grip).

[2] Using the finger is not exactly a pen based interaction but thought to be sufficiently similar to it in nature.

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 76–83, 2011.

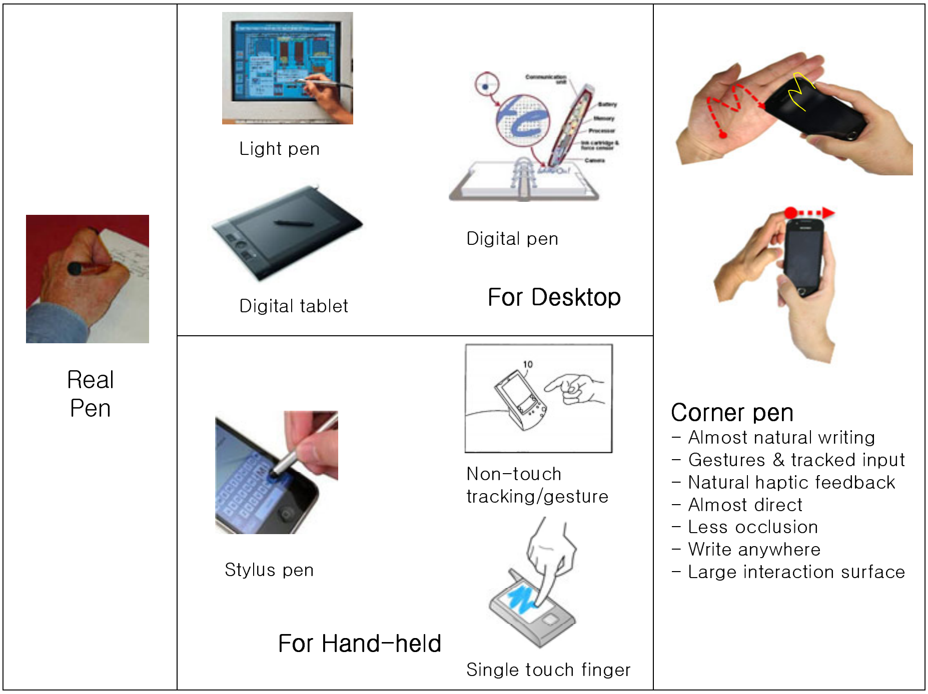


Fig. 1. Evolution of the pen-based interaction: from an actual pen to CornerPen

In this paper, we introduce the "CornerPen" in which a sensor is attached to the corner of the hand-held device (e.g. smart phone) to allow pen-based interaction using the hand-held device itself as the pen. CornerPen is a natural evolution of the pen-like input devices and has several added advantages (see Figure 1 and Section 3). Harrison and Hudson have recently introduced a very similar concept called the "Minput" in which an optical sensor was installed on the back of a small mobile device and demonstrated how simple interactions can occur through directional flick and twisting of the device [3]. Minput is particularly useful for very small mobile devices where the objects on the touchscreen are difficult to select due to their sizes relative to that of the finger tip. In our case, we are expanding the idea to a larger but still mobile device like the smart phone, *which can be held in a similar fashion to a pen (to take advantage of the precise pen-based input)*. We have implemented the proposed concept and experimentally tested its usability through two tasks, flick based icon selection (e.g. 2D menu) and free form text input, against the nominal finger driven touchscreen system.

2 Related Work

While great for direct and fast interaction, touchscreens for small mobile devices are known to have usability problems such as low precision and screen obstruction (by user's hand/finger) [1][2][3]. Thus, recent smart phones, equipped with increasingly many sensors (such as the GPS, tilt sensor, accelerometer, digital compass, camera), may offer alternative input methods. For example, many interfaces taking advantage of these sensors have appeared for recognizing gestures and tracked input in the mid-air (3D space) [4][5][6][7] but require relatively large movement of the hand-held device and thus, user is not able to see the visual output on the hand-held screen. Such input method is only suitable for non-visual applications or for controlling other external systems (e.g. interactive TV).

For relatively precise input, pen-based interface has been widely used on the desktop computing environment, for instance (e.g. especially in the design domain) either in the direct (touchscreen) or indirect (tablet) form. Similarly, for mobile devices, the stylus-touchscreen (resistive) was somewhat popular but soon replaced by the finger-touchscreen (capacitive) due to the nuisance of the stylus (or people would strive to use their finger on the resistive touchscreens).

The digital pen is a specialized device that has the same form factor as the traditional pen, but with capabilities to capture various output in digital form (e.g. text, drawings). However, it usually requires special paper and has not found widespread use. Instead, several researchers have investigated ways to use the body of the mobile device as an input medium. The use of shake or bump gestures is one such example [7][8]. Schmidt et al. developed "PhoneTouch" where phones themselves can be used for directly interacting with the objects on the touchscreen [9]. However, their system assumes the sensing to on the part of the screen or interaction surface and the interaction was devoted for objects on the system external to the phone.

3 CornerPen

3.1 Interaction Possibilities

Since the CornerPen effectively can take the same role as the stylus pen (for mobile devices) or the mouse (for desktop situation), it has a wide range of interaction applicability. The most typical interactive application would be for free form writing, being an evolution of a pen based interface. The size and form factor of today's hand-held devices make writing/manipulation almost as easy and natural as actual pen based writing/manipulation. The same task on the touch screen with the finger is much more difficult due to the relative wide touch area of the finger tip. The user is able to accomplish this task without directly using and even looking at the touch screen. While it may not replace button/key based text input, it can still find good usage for entering short memo and hand drawings.

Moreover, the input can be used for the hand-held device itself, and also for others (e.g. as a mouse for desktop computing, surface computing); that is, the CornerPen can "write" on almost any surface (e.g. palm, wall, table, forearm). Both gestures and tracked input can be realized depending on the fidelity of the sensors employed. Like the usual pen based interaction, it provides natural tactility or passive haptic feedback and even isometric input is possible with added pressure sensors. The input made (e.g. writing profile) is immediately visible through the screen and the hand make less occlusion against the hand-held screen. Alternatively, tactile or short term memory can be taken advantage of; for instance, when writing on a palm or empty surface [invisible].

Fig. 2. Few samples of interaction techniques possible with the CornerPen (from the top, gesture/character recognition, free form drawing input, tap-like input)

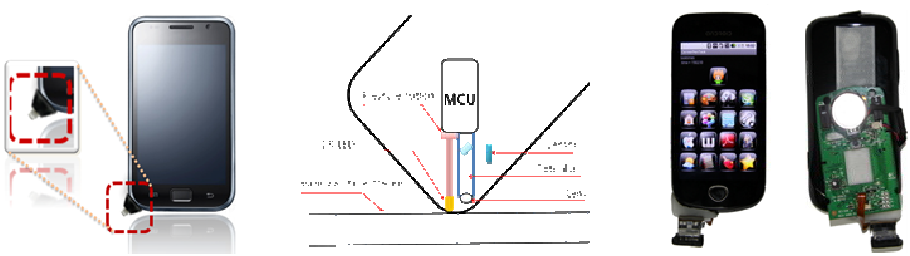


Fig. 3. The concept of the CornerPen with an optical sensor in the lower left corner of the hand-held (left). The internal architecture (middle). The actual implementation (our current version does not include the pressure sensor).

Aside from writing, as a tracking and discrete event generator, 2D gestures (e.g. simple flicker), rubber banding, drag and drop, button simulation (click/double click) are all possible in a natural fashion (Figure 2).

3.2 Implementation

In our case, we have implemented the ConerPen by adding an optical sensor (used in a standard mouse) to the lower left corner (another candidate was the upper left corner) of a smart phone. We thought making input as shown with the sensor at that location (as shown in Figure 3) was more natural and induced the phone to face the user for the display screen to be upright and visible. The optical sensor senses the infra-red light reflected by the surface and computes the relative movement of the hand-held device. The embedded MCU processes and sends (e.g. using Bluetooth communication) the sensed data to the mother hand-held device (in the actual current implementation, the sensed data is sent to a PC then back to the hand-held device).

4 Experiment

To validate the effectiveness of the proposed mobile interaction using the CornerPen, we have compared the task performance and usability between the CornerPen and the conventional touch screen based input. Two tasks were tested: (1) icon browsing (search) and selection and (2) free hand writing/drawing. The hypothesis was that the CornerPen would be superior in performance due to less occlusion (e.g. faster search) and higher precision (e.g. understandability of the drawing or writing).

Note the difference from the study of Minput [3] in that we are targeting to use CornerPen as an alternative interaction medium for general smart phone usage (rather for mere directional menu item movement in "very small" devices like a MP3 player). We hypothesized that the CornerPen would be more effective than the Finger-touch in both icon searching and making precise input due to the elimination of the occlusion problem and having a sharper contact area than the finger tip.

4.1 Experiment Design

The sole factor in the experiment was the type of the interface (ConerPen or Finger-touch) and several dependent variables were measured. The first task presented the user with several flickable (left or right) layers of icon grids (see Figure 4) from which the user was to search, find and activate an icon. Using the CornerPen, the user could flick using the corner of the hand-held device while holding the device in a way to see the whole screen without any obstruction. In case of the Finger-touch, the user would simply flick on the touchscreen to flip around the icon grid pages to search and activate the designated icon. For this task, the task completion time was measured.

The second task asked the user to enter free-form text using the two interfaces. Using the CornerPen, the user would have to hold the device and write the text as if using a pen (and see the one's own results on the hand-held screen). As for the Finger-touch, the user uses one's finger to write out the text directly on the screen (see Figure 5). For this task, self evaluated legibility (e.g. the user evaluated one's own writing in the Likert scale) was used as the dependent variable. A qualitative survey asking general preference and usability was also taken after the user tried out both treatments (1x2 Factor repeated measure).

4.2 Experiment Process

While easy enough to use, still due to the unfamiliarity, the subject was given a period of training to use the CornerPen. The subject then carried out the two tasks using the two interfaces in a balanced order. Sixteen subjects participated in the experiment (14 men and 2 women, average age: 23). The subject was first given a brief period of training to get oneself familiarized to the CornerPen interface (the touchscreen input was easy enough).

As for the first task (icon search), there were total of 64 icons divided into 4 layers (16 icons per layer) from which to find the designated icon (by flicking the layers (pages) left and right. The icon finding task was carried out for a total of 48 times in three stages (to see if there were any learning effect). The target icon and the pool of icons were all chosen and laid out in a balanced fashion.

Fig. 4. The task of browsing (searching) and selecting an icon using the Finger-touch and CornerPen

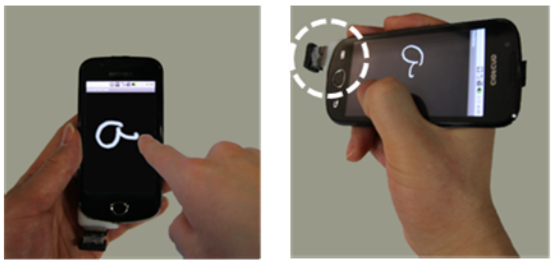


Fig. 5. The task of inputting free form text using the Finger-touch and CornerPen

As for the second task (free form writing), the subject was asked to write the seven letters ("a," "b," "c," "u," "1," "2," and "3"), 4 times for each letter. The quantitative variables were measured automatically through the task software and the survey was taken after the subject had finished testing the two interfaces.

5 Experiment Results

Figure 6 (left) shows the experiment results for the task performance between the Finger-touch and CornerPen (accumulated time for each stage, 16 times). As the graph clearly shows, the Finger-touch was far superior (with statistical significance) than the CornerPen, contrary to our hypothesis. Among many possible factors, despite prior training with the CornerPen, subjects felt much more comfortable with the usual touchscreen finger input method. The task was simple enough, despite the occasional occlusion problem with the touchscreen input, such that the CornePen did not exhibit any advantage. Further explanation is possible from the responses to the survey (later part of this section). No training effect, over the three stage trials, was found.

As for the second task, again, subjects evaluated the Finger-touch method to be much more precise with a statistical significance (Figure 6, right) due to similar reasons as from the results of the first task.

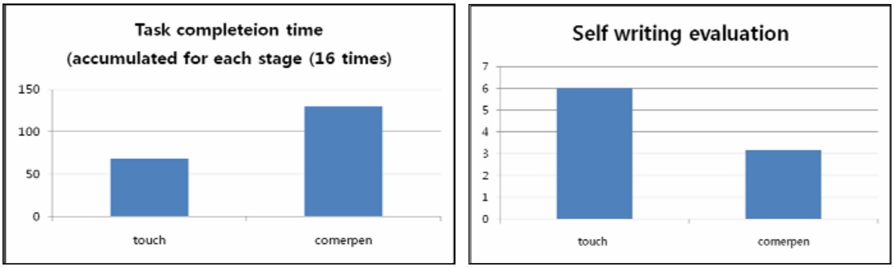


Fig. 6. The task completion time for icon search task (left) and self evaluation of the free form writing task (right) using the Finger-touch and CornerPen

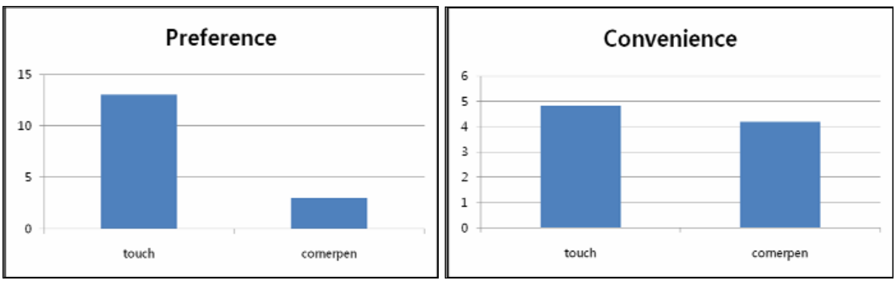


Fig. 7. User preference and ease of use responses in using the Finger-touch and CornerPen

Figure 7 shows few notable results from the survey. The subjects much preferred the familiar Finger-touch method over the CornerPen, even though convenience-wise, they were rated to be competitive (no statistically significant difference). The survey also asked of the negative influence to each input method, namely, the occlusion effect with the Finger-touch and the awkward grip needed to ensure clear visibility to the screen with the CornerPen (Figure 4 and 5). Although difficulty to directly compare, the subjects both to be problems about to the same degree.

6 Conclusion

Differently from the case of Minput [3] in which the mobile device was too small to interact with a Finger-touch, enacting the device itself as a mean for interaction was shown not to be effective when applied to a nominally sized mobile device. The naturalness and familiarity with the touchscreen outweighed the occlusion and “fat” finger tip problem. While the participants in our study reported and complained of the imperfect implementation of the CornerPen (e.g. high sensitivity and odd position of the sensor), the results indicate other solutions to these touchscreen problem might be more proper (e.g. dynamically reconfiguring the screen content according to the finger position).

In this paper, we have introduced the CornerPen in which a tracking/pressure sensor is attached to the corner of the hand-held device (e.g. smart phone) to allow the hand-held device to behave like the pen itself. While the CornerPen did not exhibit any clear advantage over the traditional Finger-touch for the typical input for the nominally sized hand-held device (e.g. smart phones), for selected tasks, it could still find good uses for, e.g. non-visual input and relative positional input for large virtual interaction space.

Acknowledgement. This research was supported in part by the Strategic Technology Lab. Program (Multimodal Entertainment Platform area) and the Core Industrial Tech. Development Program (Digital Textile based Around Body Computing area) of the Korea Ministry of Knowledge Economy (MKE).

References

1. http://www.sapdesignguild.org, Touchscreen Usability in Short (2010), http://www.sapdesignguild.org/resources/tsdesigngl/Ergo.htm#con
2. You, I., Lee, K.P.: Comparative Usability Testing between Keypad and Pen-based Device in Text Input for Small Sized Mobile Computing Devices. Korean Society of Design Science, 190–191 (2007)
3. Harrison, C., Hudson, S.: Minput: Enabling Interaction on Small Mobile Devices with High-Precision. In: Proc. of ACM CHI, pp. 1661–1664 (2010)
4. Hinckley, K., Pierce, J., Horvitz, E.: Sensing Techniques for Mobile Interaction. In: Proc. of UIST, pp. 91–100 (2000)
5. Oakley, I., O'Modhrain, M.: Tilt to Scroll: Evaluating a Motion based Vibrotactile Mobile Interface. In: Proc. of World Haptics, pp. 40–49 (2005)
6. Oakley, I., Park, J.: A Motion-based Marking Menu System. In: Extended Abstracts of ACM CHI, San Jose (2007)
7. Williamson, J., Murray-Smith, R., Hughes, S.: Shoogle: Multimodal Excitatory Interaction on Mobile Devices. In: Proc. of ACM CHI, San Jose (2007)
8. Hiner, J.: Top 25 Android Apps: Best of the Best (2010), http://www.zdnet.com/blog/btl/top-25-android-apps-the-best-of-the-best/37363
9. Schmidt, D., Chehimi, F., Rukzio, E., Gellersen, H.: PhoneTouch: A Technique for Direct Phone Interaction on Surfaces. In: Proc. of UIST (2010)

Evaluation of Continuous Practice by Mobile Learning in Nursing Practical Training

Yukie Majima[1], Yumiko Nakamura[2], Yasuko Maekawa[2], Mizuko Hiramatsu[2], Yukari Nakajima[2], Satoshi Horii[2], and Hifumi Aoyama[2]

[1] Faculty of Liberal Arts and Sciences, Osaka Prefecture University, Osaka, Japan
1-1 Gakuencho, Naka, Sakai, Osaka 599-8531, Japan
[2] School of Nursing, Osaka Prefecture University, Osaka, Japan
3-7-30 Habikino, Habikino, Osaka 583-8555, Japan
majima@las.osakafu-u.ac.jp

Abstract. Nursing students can learn many things through practical training by experiencing actual medical practice and by coming in contact with patients. Therefore practical training is an effective learning opportunity for developing the practical nursing care ability of nursing students. Moreover, at hospitals, which are important training facilities, with regard to medical safety, the use of learning tools that produce electrical waves is not possible. So, we created a learning support environment that facilitates the imagination of nursing techniques, and enables effective preparation, review, and learning at anytime and anywhere using a portable digital assistant (PDA) device for practical training. As described in this paper, we report on the outline of the educational materials named "digital nursing dictionary" that we developed and the evaluation of the practices using it.

Keywords: Mobile Learning, Nursing Practical Training, Nursing Education, Evaluation Practice, Ubiquitous.

1 Introduction

In nursing education, nursing students can learn many things through practical training by experiencing actual medical practice and by coming in contact with patients. Therefore practical training is an effective learning opportunity for developing the practical nursing care ability of nursing students. However, training facilities are scattered at various locations, and do not necessarily provide the environment required for learning (library, internet, or IT equipment). Furthermore, depending on where practical training takes place, there are also students who spend a lot time commuting, and thus, it is important to provide an environment where effective learning is possible during practical training. Moreover, at hospitals, which are important training facilities, with regard to medical safety, the use of learning tools that produce electrical waves is not possible. Therefore, we created a learning support environment that facilitates the imagination of nursing techniques, and enables effective preparation, review, and learning at anytime and anywhere using a portable digital assistant (PDA) device for practical training. We call this PDA "digital nursing dictionary". This study purpose is to evaluate the mobile learning methods like using the PDA during the three-year period of nursing practical education.

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 84–91, 2011.

2 Background

To respond to that social need for human resources, we believe that it is necessary to improve the education and instruction of nursing more qualitatively, and to provide nursing students with a new environment in which they can study efficiently.

Among all educational activities, the most effective way for students to acquire actual nursing capabilities is to have the learner join on-site practice training, which is a required part of the nursing education curriculum. Through on-site practice, students can learn comprehensively by communicating directly to patients or those who require nursing care. However, such practicing facilities are spread out in various locations. For that reason, the situation available on site is not necessarily as good a study environment as that available on campus. The campus might provide facilities such as libraries, Internet capability, and IT equipment. By making e-Learning available at every on-site practice location, we believe that we can not only improve the learning environment situation, but also help to raise the quality level of the learning itself.

So we developed the nursing e-learning environment in 2005 as shown in the Figure 1. This e-learning environment consists of nursing training materials with examples for each specialized nursing area based on the training frames for e-Learning of nursing[1]-[4] and "digital nursing dictionary"[5]. Nursing students can download examples of nursing from the nursing training materials server machine. Each student is allowed to compile a "digital nursing dictionary" by downloading the necessary sub-training materials before the on-site practice starts. A "digital nursing dictionary" is shown in Figure 2. In the other mobile method, Maag[6] introduced an emerging technology using by podcasting in nursing education and the PDA e-portfolio tool[7] was developed by Bernard et al.

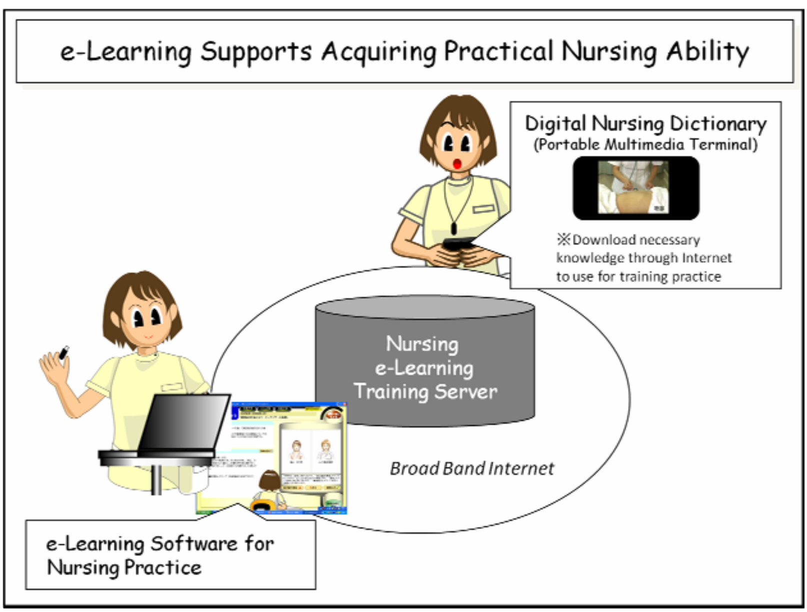


Fig. 1. Nursing e-learning environment in Osaka Prefecture University

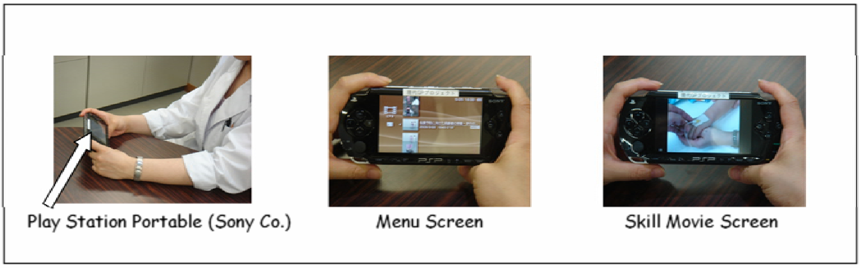


Fig. 2. Digital nursing dictionary

3 Methods

During the three years (2007, 2008, and 2009) of practical training at the school of nursing in our university, we prepared portable multimedia terminals (PSP, Play Station Portable; Sony Corp.) loaded with training material contents (digital nursing dictionary) prepared for the associated nursing practice areas, and lent them each to students who agreed to monitor them. Subsequently, we requested that they use them during the entire practice period. We requested them to give us opinions regarding the operability of the training materials and other information when they had to return the PDA.

Practical training for continuous practice was of three types: practical training II (second-year students), basic practical training (third-year students), and applied practical training (fourth-year students). During the three years of continuous practice by mobile learning, students could borrow this dictionary multiple times.

When returning the digital nursing dictionary after training was completed, a self-administered anonymous survey was conducted. The survey content consisted of free comments on the usage of the digital nursing dictionary and the usefulness of its content. The protocol of the survey was approved by the research ethics committee of the university. The survey period ranged from April, 2007 to March, 2010. We compared the interannual data for each practical training.

The ethical affairs committee for research of the school to which all of our study group members belong approved the study project we conducted.

4 Results

4.1 Actual Borrowing of Digital Nursing Dictionaries

Table 1 shows the borrowing rate of the digital nursing dictionaries for each school year from 2007 to 2009. In 2007, the borrowing rate of the dictionaries for applied practical training (subjects in the first term of the fourth grade) was 26.2%, which was low. This may be because 2007 was the first year continuous practice by mobile learning was conducted, and due to the schedule of the preparation of the content, the dictionary was borrowed by the students after practical training had begun (May to June). Therefore, in the same year for basic practical training (subjects in the second term of third year), the date of issue was set immediately after the orientation on practical training, and lending of this dictionary was arranged in a way that it could be borrowed at places where students gathered when they felt motivated.

Thus, from the second term of practical training in 2007, as a result, >90% of students wished to borrow this dictionary. Furthermore, in 2007 there were many free comments such as “there is not enough time for getting used to this new teaching material and to become able to use it.” Thus, setting the date of issue at the earliest possible time caused the borrowing rate to increase from 60% to 90%.

Table 1. The borrowing rate of the digital nursing dictionaries for each school year and practical training from 2007 to 2009

year	First Practical Training 2nd grade (2 weeks)	Basic Practical Training 3rd grade (4 months)	Applied Practical Training 4th grade (3 months)	total
2007	78.4% [98/125]	93.3% [111/119]	26.2% [33/126]	65.4% [242/370]
2008	88.2% [105/118]	91.5% [119/130]	65.4% [85/130]	81.7% [309/378]
2009	77.1% [91/118]	76.9% [90/117]	87.3%[117/134]	80.8% [298/369]

[borrowing student number/all student number]

4.2 Usage of the Digital Nursing Dictionary in the Different Types of Practical Training

Regarding first practical training (second-year students) (Figure 3), there were annual variations in the percentage of students who responded “I often used it” or “I used it”. The learning goal of this practical training was to investigate the requirements for nursing, such as how to communicate well with and gain a comprehensive understanding of the patient. Due to diversity of the patients that students have to deal with, the teaching material content of the digital nursing dictionary is difficult to use in cases that do not match the provided examples.

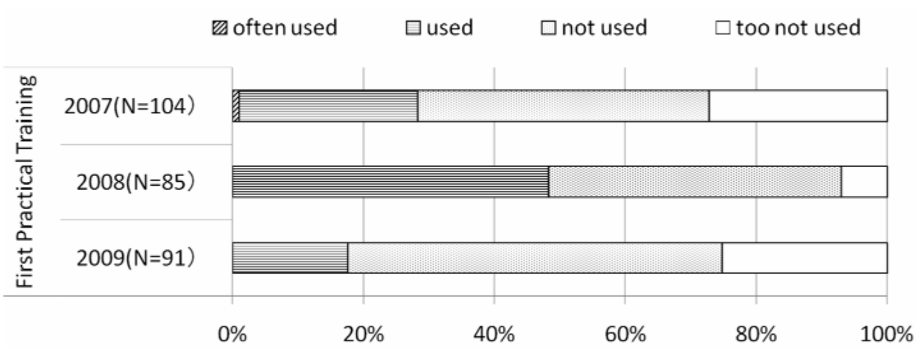


Fig. 3. Usage of the digital nursing dictionary in the first practical training

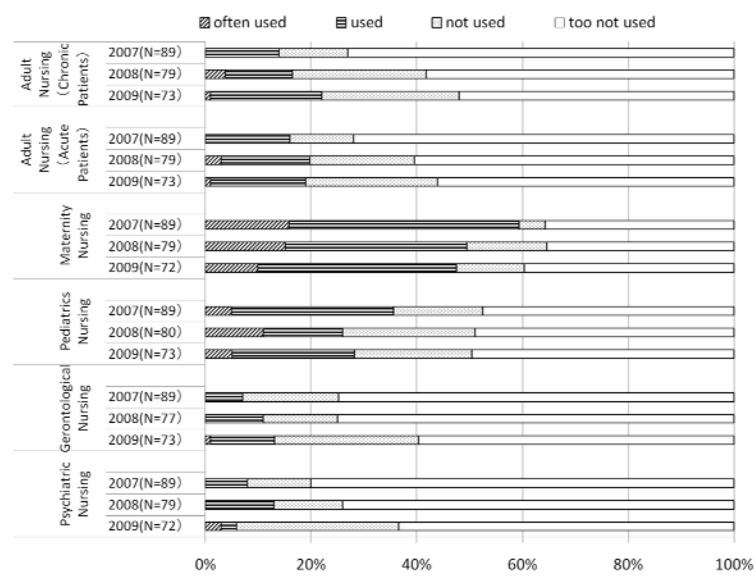


Fig. 4. Usage of the digital nursing dictionary in the basic practical training

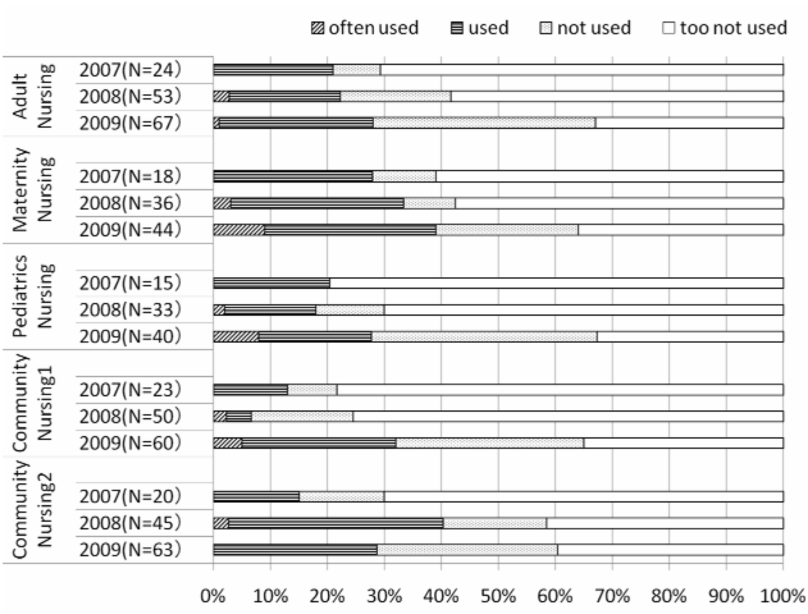


Fig. 5. Usage of the digital nursing dictionary in the applied practical training

On the other hand, for basic practical training (Figure 4) and applied practical training (Figure 5), the teaching material focuses on practical training in a subject-dependent manner; thus, necessary nursing techniques and knowledge are narrowed down. The rate of utilization was in particular high in the fields of maternity and pediatrics because the content featured in these fields is strongly related to the content necessary for the practical training. In addition, the rate of its utilization has been increasing over the years for adult nursing (chronic patients) in basic practical training and community nursing in applied practical training. For community, nursing images on nursing techniques, which are included in the digital nursing dictionary, were developed and used as preliminary study tasks for practice at the nursing school prior to nursing practical training.

4.3 Usefulness of the Digital Nursing Dictionary Contents

This dictionary included nursing technique images, and knowledge cards. The results of their respective usefulness in 2008 and 2009 for practical training are shown in Figure 6. Approximately >60% of the students responded that the images on nursing techniques were "very helpful" or "helpful" for the entire practical training, and approximately >50% responded that knowledge cards were "very helpful" or "helpful" for the entire practical training. Furthermore, regarding the evaluation of the use of the dictionary, >40% of the students responded that "it was helpful that they could study at home," "videos were easy to understand," and "images were clear and easy to understand" during the entire practical training.

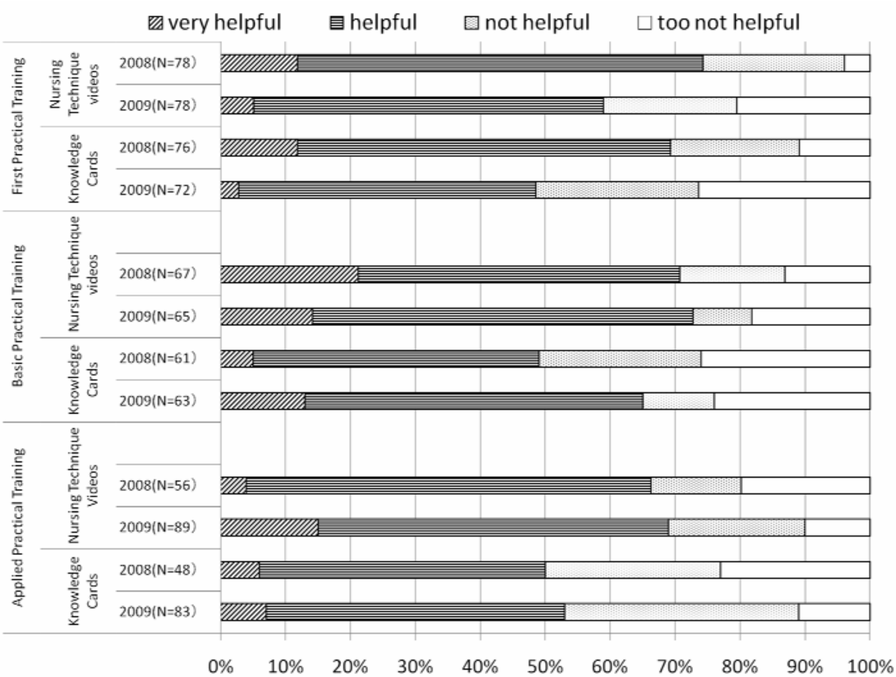


Fig. 6. Usefulness contents of the nursing dictionary

5 Discussion

After examination of the borrowing rate throughout basic and applied practical training in 2007 and 2008 (Table1), we found that borrowing rate decreased with regard to applied practical training. This indicates that among the various learning methods students choose learning methods that suit their own needs when borrowing teaching material. The results of this study showed that the reason why the borrowing rate was >60% for nursing practical training was that students want teaching material that enables them to imagine nursing techniques, which they then can continuously practice by themselves. We believe that this method in which teaching material was provided in easily PDAs that can be used at home and/or during commuting and that provide clear images was helpful. Furthermore, it was suggested that having the opportunity to actually use this dictionary prior to the practical training at school leads to more effective use in the course of continuous practice. We believe that it is necessary to actively increase the opportunities for students to use such learning materials, and at the same time to make its content widely known and promote it among the teaching.

6 Conclusion

In this paper, we reported on the outline of the educational materials named "digital nursing dictionary" that we developed and the evaluation of the continuous practices using it during the three years (2007, 2008 and 2009). We found the usages and the usefulness of the digital nursing dictionary for nursing students in the practical training at hospitals.

Finally, this paper suggested that having the opportunity to actually use this digital nursing dictionary prior to the practical training at school leads to more effective use in the course of our continuous practice.

We believe that it is necessary to actively increase the opportunities for students to use such learning materials

References

1. Yukie, M., Yoichiro, S.: Development of E-learning for Problem solving Approach of Nursing Students. Studies in Health Technology and Informatics 122, 919 (2006)
2. Kazuhisa, S., Yukie, M., Yoichiro, S.: Nursing task ontology based learning system redesign for enabling adaptive on-demand learning environment. In: Proceedings of the 3rd WSEAS/IASME International Conference on Engineering Education, Greece, pp. 114–119 (2006)
3. Kazuhisa, S., Yukie, M., Yoichiro, S.: Towards Building Adaptive On-Demand Learning Environment Based on Nursing Ontology. WSEAS Transaction on Advantages in Engineering Education, 563–570 (2006)
4. Yukie, M., Yoichiro, S., Kazuhisa, S.: Framework for Problem-Solving Based Learning in Nursing Domain -An Experimental Study-. In: Learning by Effective Utilization of Technologies: Facilitating Intercultural Understanding, pp. 625–628 (2006)

5. Yukie, M., Yumiko, N., Yasuko, M., Yoichiro, S.: Development of ubiquitous on demand study support environment for nursing students. In: Federrath, H. (ed.) Designing Privacy Enhancing Technologies. LNCS, vol. 5613, pp. 149–155. Springer, Heidelberg (2009)
6. Margaret, M.: iPod, uPod? An emerging mobile learning tool in nursing education and students' satisfaction. In: Proceedings of the 23rd Annual Ascilite Conference, pp. 483–492 (2006)
7. Bernard, M.G., Cathryn, J.: A mobile clinical e-portfolio for nursing and medical students, using wireless personal digital assistants (PDAs). Nurse Education Today 26, 647–654 (2006)

XML in Formal Specification, Verification and Generation of Mobile HCI

Ines Riahi, Meriem Riahi, and Faouzi Moussa

Laboratoire Cristal, Ecole Nationale des Sciences de l'Informatique de Tunis, rue des Entrepreneurs Charguia II, 2035 Tunis-Carthage, Tunisia
ines.riahi@yahoo.fr, Meriem.Riahi@insat.rnu.tn,
faouzimoussa@gmail.com

Abstract. Our work is carried out in the framework of a global approach for Human-Computer Interaction design and automatic generation. This approach is based on a formal modeling of the Human-Computer Interaction. We propose a Model Based Design Approach (MBDA). We are concerned with identifying the user tasks and requirements and further with the automatic graphical interface validation and generation. Therefore we use Petri Nets. Indeed, the Petri Nets are very efficient in formal modeling of HCI. Our research focuses on mobile HCI. It aims to analyze the ubiquitous environment using ontology described in OWL2 standard. We face difficulties in modeling ontology in XML using Petri Nets. Thus, it becomes necessary to adopt approaches for manipulation of Petri nets via XML as PNML or XML Nets.

Keywords: Human-Computer Interaction (HCI), Model Based Design Approach (MBDA), Ubiquitous environment, Petri Nets, Ontology, OWL2, XML, XML Nets, PNML.

1 Introduction

Our research field is Human-Computer Interaction (HCI) and it concerns especially mobile HCI. We intend to elaborate a global approach for formal specification and generation of mobile interface. In fact, ubiquitous environment is characterized by an important rate of information received from different sources. To represent this knowledge, ontology- described in the standard OWL2 (Ontology Web Language V2) based on XML (Extensible Markup Language) - is used. For modeling the Human-Computer Interaction, we use the Petri Nets. We face difficulties in modeling the ontology using Petri Nets and how to express the interaction between an ontology represented by OWL2 and Petri nets formalized by XML. Thus, it becomes necessary to adopt approaches for manipulating Petri Nets via XML. PNML and XML Nets are proposed to deal with this problem.

Our paper focuses on the use of XML in formal specification (with Petri Nets), verification and generation of mobile graphical interfaces. The first part of this paper presents a global approach for formal specification and generation of mobile graphical interfaces. Then, a bibliographic review of the methods XML Nets and PNML is presented and discussed. Finally, we propose our approach based on the profit of using both XML and Petri Nets.

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 92–100, 2011.

2 Approach for Formal Specification, Verification and Generation of Mobile HCI

The effervescence of new technologies and means of mobile communication in recent years has inspired developers to bring mobile devices into their applications giving rise to ubiquitous computing. This new environment integrates mobile devices with different hardware and software capabilities.

The adaptation of human-computer interfaces to an ubiquitous environment consists in producing a number of interfaces that can be dynamically adapted to different types of mobile devices while respecting these ergonomic properties. Most of the works done on adapting mobile HCI are based on models describing the interaction between human and machine. [1]

The mobile HCI for PDA, mobile phones or tablets are characterized mainly by small screens, usually limited treatment capacity, backup information and the availability of random network [2]. All these constraints must be taken into account when modeling the HCI Mobile.

Until now, most of the works has focused mainly on technological aspects of mobile terminals or on the problems of assessing the usability of mobile devices without considering a crucial point which is modeling mobile HCI. These devices must continue to be functional in any environment in which they are used. This requires the definition and the integration of the concept of mobility in a model based approach.

In this context, we propose an approach that is made up of five steps as shown in figure 1:

1. The first step consists in analyzing the whole Human-Machine System in terms of the system (in different contexts and environments) [3], the interaction and the user's tasks. The modeling of the System's behavior becomes possible. It expresses the interaction of the User with the Graphical Interface. This analysis is carried out using Petri Nets modeled in XML. This is discussed further, in section 4.
2. The second step is the achievement of the deduction of the user requirements. In fact, the ubiquitous environment analysis and its modeling provide the set of user requirements in accordance with each functional context.
3. The third step ensures the identification of the interface objects according to the user requirements. Once the interface objects are identified, this step consists in specifying the interface in terms of displays, graphical objects and dialogue.
4. In the next step, we take advantage of the formal technique used for interface specification to verify critical properties of the generated specifications. Indeed, formal techniques are specially recommended because they allow the designers to proceed to the validation (even mathematical) of the UI (User Interface) before going on to their actual creation and implementation.
5. The last step in this approach is devoted to the automatic interface generation.

We focus, in this paper, on the first step: Human-Machine System Analysis in an ubiquitous environment. First, we study the environment to detect the different characteristics and functional contexts. Then, we identify, for each context, the user tasks. The task analysis is, afterward, carried out.

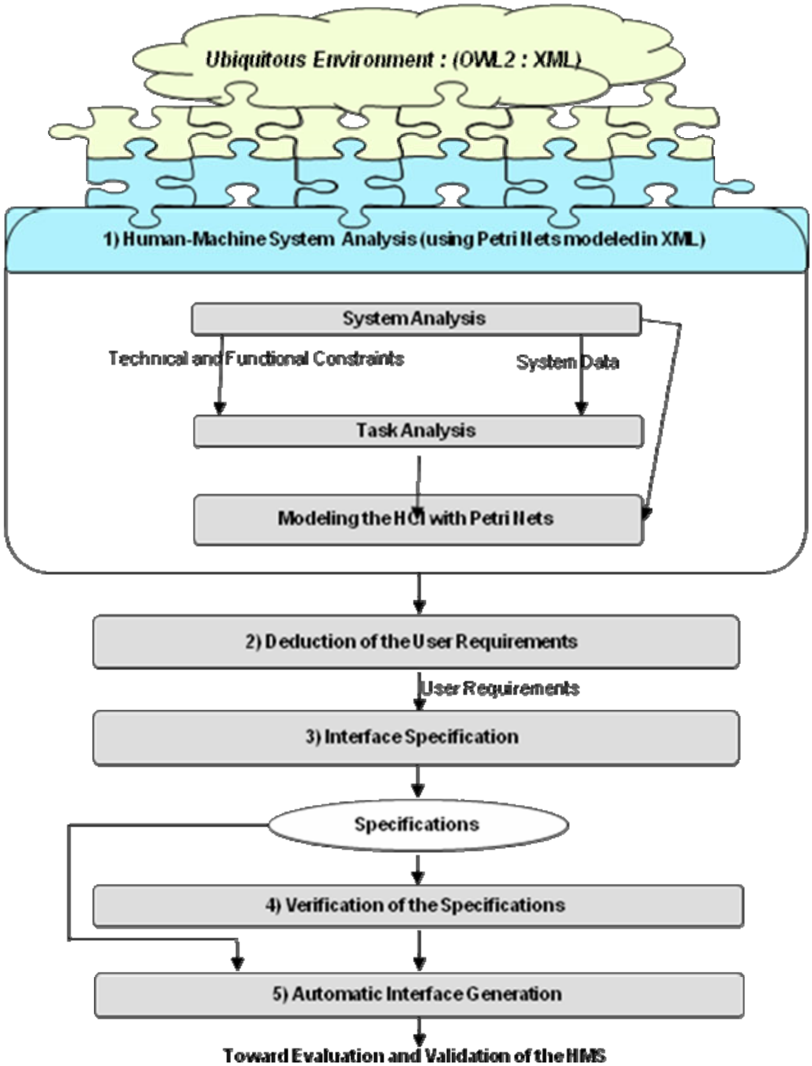


Fig. 1. Proposed Approach for HCI Automatic Generation [4]

This analysis allows the definition of the operator's actions according to different contexts. The parameters of these actions will constitute, therefore, the user requirements. Thus, this analysis will provide a document containing the different variables necessary for the mobile HCI modeling. This document is described in OWL2.

Once this is done, we proceed to the modeling of the Human-Computer Interaction in terms of operator's actions. This modeling takes into account the different contexts and evolution of the environment. This modeling must be performed with a formal technique. This allows validating certain properties of the interface as stability and absence of blocking and conflict. Petri nets are proposed for that. The difficulty, here, is how to combine Petri nets with XML documents obtained by the description in OWL2.

The section below presents a bibliographic study about these formalisms.

3 XML and Petri Nets

3.1 XML NETS

XML nets are a new variant of high-level Petri nets. They are a formal graphical modeling language which models the flow of XML documents. The choice of Petri nets was obvious. Petri nets have the advantages of the processes' graphical representation with a formal semantic modeling of the behavior. XML nets are based on GXSL, a graphical XML schema definition language, and the corresponding XML document manipulation language XManiLa.

Graphical XML schema definition language (GXSL): is a graphical language for XML schema definition for the design of XML document types. It consists of a set of markup declarations of different types: element, attribute list, entity or notation declaration. GXSL is based on DTD (Document Type Definition) that can be derived from XML schema. Instead of creating a completely new graphical modeling language for XML document types, we rely on well known data modeling concepts such as the Unified Modeling Language (UML) [5].

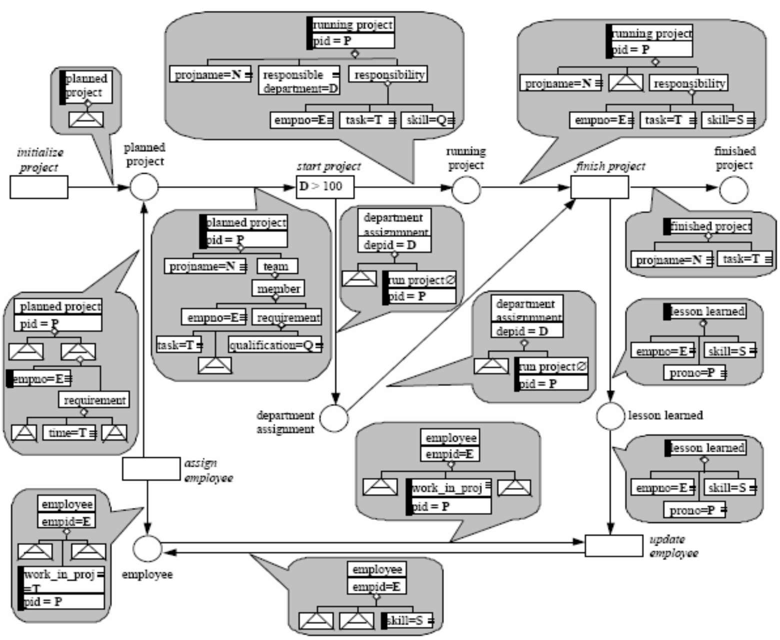


Fig. 2. Example of XML Nets [5]

The presented version in [5] of GXSL is DTD-based, i.e. a DTD can be unambiguously derived from the XML schema diagram XSD. The full GXSL provides concepts for graphically modeling data type-based XML schemas with enhanced features like an additional kind of element aggregation, element substitution, element and association references, and finally the concept of keys and foreign keys.

XML document manipulation language (XManiLa): XManiLa is the GXSL based XML document manipulation language, which can use GXSL with some extensions also for querying and manipulating XML documents. In order to enable also a content based query, we allow for assigning constants or variables to an element or an attribute. XManiLa is not only suited for document retrieval, but also for insertion and deletion (and thereby also for updating). These operations either concern a whole document or elements on lower hierarchy levels. [5]

The combination of GXSL with XMANILA gives XML Nets as shown in figure 2. The static components of XML nets (i.e. the places of the Petri net) are typed by XML schema diagrams, each of them representing a DTD. Places can be interpreted as containers for XML documents which are valid for the corresponding DTD. The flow of XML documents is defined by the occurrences of transitions.

3.2 Petri Net Markup Languages (PNML)

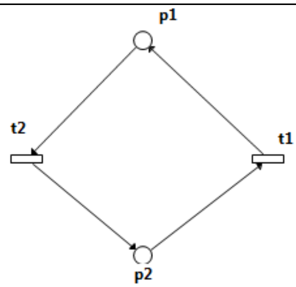
The Petri Net Markup Language (PNML) is an XML-based interchange format for Petri nets. The design of PNML was governed by the following principles:

- Readability: The format must be readable and editable with any text editor,
- Universality: The format should be able to represent any kind of Petri nets with its possible extensions,
- Mutuality: The format should allow us to extract as much information as possible from a Petri net. Therefore, the format must extract the common principles and the common notations of Petri nets [6].

The use of XML guarantees the readability of the format. Universality can be guaranteed by labeling net objects and the net itself. The legal labels, their possible values, and the possible combination of values are defined by a Petri Net Type Definition (PNTD). Mutuality can be guaranteed by conventions, which are a set of standardized labels [6].

The main idea of PNML is that any kind of Petri net can be considered to be a labeled graph. In particular, all information that is specific to a particular kind of Petri net can be captured in labels.

A file that meets the requirements of PNML is called a Petri net file; it may contain several Petri nets. Each Petri net consists of objects, which, basically, represent the graph structure of the Petri net. Each object within a Petri net file has a unique identifier, which can be used to refer to this object. In basic PNML, an object is a place, a transition or an arc. For convenience, a place or a transition is called a node. In order to assign further meaning to an object, each object may have labels. Typically, a label represents the name of a node, the initial marking of a place, the guard of a transition, or the inscription of an arc. Two kinds of labels are distinguished: annotations and attributes. An annotation comprises information that is typically displayed as text near the corresponding object. In contrast, an attribute specifies a graphical property of an object such as color, style, form, or line thickness. Each object and each annotation is equipped with graphical information. For some tools, it might be necessary to store tool specific information, which is not supposed to be used by other tools. In order to store this information, each object and each label may be equipped with such tool specific information. Its format depends on the tool and is not specified by PNML [7].




```
<?xml version='1.0' encoding='UTF-8'?> <model ….
<nodes> <node nodetype='place' …
<attributes><attribute name='name' xposition='296'
yposition='218'>p2</attribute>          <attribute
name='marking' xposition='0'
yposition='0'>0</attribute></attributes>    </node> <node
nodetype='place' … <attributes><attribute name='name'
xposition='315' yposition='19'>p1</attribute><attribute
name='marking' xposition='0'
yposition='0'>0</attribute></attributes> </node> <node
nodetype='transition' … <attributes> <attribute
name='name' xposition='201'
yposition='106'>t2</attribute></attributes> </node> <node
nodetype='transition' … <attributes> <attribute
name='name' xposition='406'
yposition='114'>t1</attribute> </attributes> </node>
</nodes> <arcs> <arc arctype='arc' id='7' startid='6'
endid='4'> </arc> <arc arctype='arc' id='8' startid='4'
endid='5'> </arc> <arc arctype='arc' id='9' startid='5'
endid='3'> </arc> … </arcs> </model>
```

Fig. 3. Example of PNML

In order to guarantee compatibility among different Petri net types, the Conventions Document comprises the definitions of all standard labels; technically, the Conventions Document consists of a sequence of more or less independent label definitions. Each label definition in the Conventions Document must be assigned a unique name. For each label, the Conventions Document gives a reference to its meaning and states in which PNML elements it may occur. The conventions Document guarantee that the same labels have the same meanings in all types of Petri nets which allow the exchange between them [8]. The above example (figure 3) shows the description of a Petri nets with two places and two transitions.

3.3 Discussion

As a variant of high-level Petri nets as shown in figure 2, XML nets have formal semantics, graphical nature, and the strength in exchanging XML-based structured data. They are very suitable for Web Service Composition WSC because messages can be modeled and manipulated as place tokens for message passing, and the labels in arcs can be used to model constraints for WS discovery and selection. Using XML nets for WSC can thus improve WSC models and increase their dynamics [9]. In addition, XML Nets support inter-organizational business processes [10] and the supply chain management [11].

XML Nets have additional advantages in the description of process objects and inter-organizational exchange of standardized structured data (e.g. XML documents) [10]. However, XML Nets cannot verify the validity of Petri Nets. We may, also, get complex XML schemas for a complex system.

The Petri Net Markup Language (PNML) is an XML-based interchange format for Petri nets. In order to support different versions of Petri nets and, in particular, future versions of Petri nets, PNML allows the definition of Petri net types. Due to this flexibility, PNML is a starting point for a standard interchange format for Petri nets.

Another important issue for interchange formats is the size of real world systems. Typically, real world systems are too large to be models. PNML Modular, extension of PNML, introduces the concept of modules. Thus, a system can be built recursively from module instances (sub-Petri Nets) [6].

However, this interchange format does not guarantee the accuracy and validity of Petri nets obtained on sub-Petri Nets used recursively [12]. In addition, neither XML Nets nor PNML Nets have been used in the field of Human-Computer Interaction specification.

4 Proposed Approach

We are interested in our research work, in mobile HCI. Ontology is proposed to represent and formulate the knowledge of an ubiquitous environments. It is a set of structured concepts and relationships modeling the knowledge and providing a sense to each information [13]. OWL2 is proposed for this fact. Indeed, this web ontology language is based on a dialect of XML.

So we use XML as a mediator language allowing the analysis and the modeling of the ontology. Our contribution is to elaborate a model of a HCI based on Petri Nets using XML documents. XML guarantees the exchange of the ubiquitous environment and the mobile HCI specification as shown in figure 4.

In fact, Petri nets will specify the behavior and the treatment that will undergo the data extracted from the ontology. This data will “feed” the Petri nets gradually along the evolution of the environment.

The first step of this approach consists in translating Petri nets in XML documents. For each transition, we associate the useful concepts of the ontology. These concepts written in an XML document are used to execute the adequate tasks.

Thus, the result will be an XML document allowing the deduction of the user requirements and the generation of the user interface.

This approach can be appropriately applied to the example the car speed controlling.

This application is intended to control the speed of a driver according to her environment. This management is produced through a mobile HCI introduced in a mobile device.

Whatever its nature: PDA, GPS or a mobile phone, its role is to capture the nature of the environment and adjust the speed of the driver according to these data. These are defined using ontology written in OWL2 as we have specified in the previous section.

Then the Human-Machine Interaction while the operator is performing these tasks will be adapted using one of the two approaches presented in the third section.

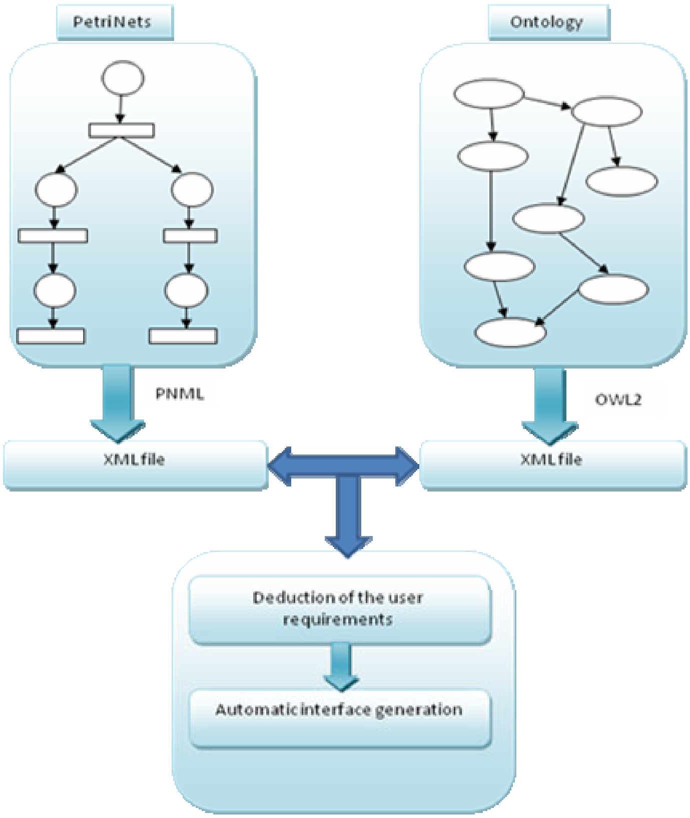


Fig. 4. Proposed approach

OWL data will feed into our Petri nets with XML based on the treatment chosen by the driver. Thus, from a technical point of view, we have two XML files. One representing our ontology which contains the description of our environment and the other contains the modeling of the Human-Computer System. Further, we will present an algorithm that ensures the interaction between these two XML files.

The output of this algorithm is an XML file that will allow the automatic generation of mobile HCI. This file will take into account the characteristics of mobile HCI contained in the first section.

The analysis and the modeling of ubiquitous environments are difficult and complex.

PNML Modular, extension of PNML, presented in section 3, offers the possibility to reduce the complexity by dividing a complex system in a set of simpler modules. These modules can be, afterward, used in a recursive manner.

In our previous works, we have focused on "how to make composition" while modeling the HCI with Petri Nets. According to these composition rules, we guarantee the validity of the Petri Net's critical properties. With verified properties, it becomes possible to validate the user interface. We combined these concepts with XML notation in order to propose a new extension of the standard PNML. These works will be presented in the near future.

5 Conclusion

In this paper we use the Petri nets with XML to model the system, the user's task and the interaction. A user's task is composed of a well-organized set of elementary actions. A global model is then constructed integrating the different evolutions of the functioning system's states and the associated user's tasks. The purpose of this work is to be able to deal with ontology while using a Petri Nets based modeling approach. This allows us to better link the two steps (i); the description of the ubiquitous environment using ontology (OWL2), (ii) and the formal modeling of the mobile Human-Computer Interaction. Further, we will take advantage of the formal technique used to verify critical properties of the generated specifications of the user interface.

References

1. Chaari, T.: Adaptation d'applications pervasives dans des environnements multi-contextes. Thèse de doctorat, L'institut national des sciences appliquées de Lyon (2007)
2. Calvet-Inglad, G.: Etude empirique de l'interaction multimodale mobile en situation naturelle. Thèse de doctorat, Université de Toulouse (2008)
3. Riahi, M.: Contribution à l'élaboration d'une méthodologie de spécification, de vérification et de génération semi-automatique d'interfaces homme-machine: Application à l'outil Ergo-Conceptor+. Thèse de doctorat, Université de Valenciennes et du Hainaut-Cambrésis (2004)
4. Moussa, F.: Vers une méthodologie globale de conception et de génération semi- automatique des IHM pour les systèmes industriels. Habilitation Universitaire en Informatique, Faculté des Sciences de Tunis (2005)
5. Lenz, K., Oberweis, A.: Inter-Organizational Business Process Management with XML Nets. In: Ehrig, H., Reisig, W., Rozenberg, G., Weber, H. (eds.) Petri Net Technology for Communication-Based Systems. LNCS, vol. 2472, pp. 243–263. Springer, Heidelberg (2003)
6. Weber, M., Kindler, E.: The Petri Net Markup Language. In: Ehrig, H., Reisig, W., Rozenberg, G., Weber, H. (eds.) Petri Net Technology for Communication-Based Systems. LNCS, vol. 2472, pp. 124–144. Springer, Heidelberg (2003)
7. Hillah, L.M., Kindler, E., Kordon, F., Petrucci, L., Trèves, N.: A primer on the Petri Net Markup Language and ISO/IEC 15909-2. Petri Net Newsletter (2009)
8. Billington, J., Christensen, S., van Hee, K.M., Kindler, E., Kummer, O., Petrucci, L., Post, R., Stehno, C., Weber, M.: The Petri Net Markup Language: Concepts, Technology, and Tools. In: van der Aalst, W.M.P., Best, E. (eds.) ICATPN 2003. LNCS, vol. 2679, pp. 483–505. Springer, Heidelberg (2003)
9. Che, H., Li, Y., Oberweis, A., Stucky, W.: Web Service Composition Based on XML Nets. In: Proceedings of the 42nd Hawaii International Conference on System Sciences (2009)
10. Che, H., Mevius, M., Ju, Y., Stucky, W., Trunko, R.: A Method for Inter-organizational Business Process Management. In: Proceedings of the IEEE International Conference on Automation and Logistics, Jinan, China (August 18-21, 2007)
11. Che, H., Stucky, W., Ju, Y.: Using XML Nets and Grid Services to support SCOR. In: Proceedings of the Seventh International Conference on Machine Learning and Cybernetics, Kunming (July 12-15, 2008)
12. Vidal, J., Lama, M., Bugarin, A.: A High-level Petri Net Ontology Compatible with PNML. Petri Net Newsletter (2006)
13. http://www.technoscience.net/?onglet=glossaire&definition=324

An Efficient Document Browsing Method with Floating Diagram Window on Mobile Device

Yu Shibuya, Kazunobu Nagata, and Kazuyoshi Murata

Kyoto Institute of Technology,
Matsugasaki, Sakyo-ku, Kyoto 6068585, Japan
shibuya@kit.ac.jp

Abstract. This paper proposes an efficient method to browse the document including diagrams on mobile devices equipped with touch screen. With the method, the user can pull the referenced diagram close to the referrer sentence and browse the pulled-in diagram and the document simultaneously. The pulled-in diagram is displayed in the floating window overlapping the document. Both the pulled-in diagram and the underlying document are able to be scrolled and zoomed in/out by the user independently of each other. From the experiment, it is found that the proposed method is efficient for operation on mobile devices and the method is viewed very positively by the participants.

Keywords: mobile interaction, document browsing, touch screen, human interface.

1 Introduction

It is very common today to use mobile devices like cell phones or PDAs (Personal Digital Assistant) for browsing web pages or documents. Furthermore, recent mobile devices are equipped with a high-resolution touch screen so that the user can scroll or zoom in/out with intuitive finger gesture on the screen.

While users browse a document including some diagrams, they often want to see both referrer sentence and its referencing diagram simultaneously. However, it is difficult to do so with mobile devices because of their small screen size. On such devices, users have to scroll the document to see the desired sentence or diagram. Furthermore, if they want to see the sentence and the diagram mutually, they have to scroll so often that it takes long time to browse the document. It is also time consuming operation to zoom in/out the contents to get the suitable view magnification.

This paper proposes an efficient method to browse the document including diagrams on mobile devices equipped with touch screen.

2 Proposed Method

Our proposed method provides a floating diagram window overlapping the document. With our proposed method, the user can pull the referenced diagram close to the referrer sentence with following manner. Firstly, he/she touches a referrer word, such as “figure x” or “table y”, on the screen, and then a string-like graphical line is drawn

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 101–106, 2011.

between the word and the referenced diagram. If he/she pulls the string with his/her swipe-in finger gesture, the diagram is pulled into the screen and shown in the floating diagram window. Fig. 1 shows how to pull a desired diagram into the screen.

The pulled-in floating diagram window overlaps the document. The user can touch the window's border and drag the window to desired position. He/she can also scroll or zoom in/out the diagram in the window. He/she can scroll it with swipe finger gesture and zoom it with pinch finger gesture. At the same time, with the same manner, he/she can also scroll or zoom in/out the underlying document. Both the pulled-in diagram and the underlying document are able to be scrolled and zoomed in/out by him/her independently of each other. As a result, with the proposed method, the user can browse the document and the diagram simultaneously on the small screen of mobile devices.

When the pulled-in diagram becomes to be unnecessary for the user, it can be thrown away from the screen with the swipe-out finger gesture with touching the border of the diagram window.

In order to implement the proposed method, there must be a link between the referrer word and its referencing diagram. This paper focuses on HTML documents which already have such links. It might be our future work to propose and implement a method for making a link between the referrer word and its referencing diagram in other type of document.

(a) (b)

Fig. 1. With the proposed method, (a) a user can touch a referrer word on the screen and (b) pull the referenced diagram into the screen with swipe finger gesture

3 Experimental Evaluation

The purpose of experiment is to examine the performance of the proposed method to browse the document including some diagrams on mobile devices. The proposed method is compared with a conventional method described below.

3.1 Compared Conventional Method

With a conventional method for browsing the document on mobile devices with touch screen, users can simply scroll the document with swipe finger gesture and zoom in/out it with pinch finger gesture. Our proposed method is compared with this conventional method experimentally.

3.2 Prototype System

Both proposed and conventional methods are implemented on the Apple's iPod touch (3rd generation, 8GB memory). Its screen resolution is 320 by 480 pixels but top of 20 pixels are used as the status bar. Thus the resolution for displaying the document is 320 by 460 pixels.

In the prototype system of the proposed method, the resolution of floating diagram window is 320 by 230 pixels and the width of each border is 30 pixels. Thus the resolution for displaying the diagram is 260 by 170 pixels. The width of diagram display area is a little bit narrower than that of document display area. The height of diagram display area, on the other hand, is less than a half of the height of document display area so that the user can read the underlying document around the floating diagram window.

3.3 Procedure

Ten participants were recruited from our university. They were undergraduate and graduate students (nine male, one female). A repeated measurements within-subject design is used for the experiment.

In the experiment, each participant was asked to hold the mobile device and to tap the start button on the screen. Then a HTML document with some figures was shown on that screen. The participant was asked to find two questions in the document and to answer orally one by one. In order to answer each question, he/she must see the figure referred in the question. After he/she answered both questions, he/she pressed the hardware stop button below the screen.

Task completion time, the elapsed time from pressing start button to pressing stop button, was measured. Subjective evaluation was also done with a questionnaire after each task.

4 Results and Discussions

Through the experiment, all participants answered all questions correctly. This means that both methods allowed users to browse the document without serious problems.

4.1 Task Completion Time

From the experiment, as shown in Fig. 2, it is found that the task completion time with the proposed method was 28.5 seconds and was slightly shorter than that of conventional method (30.5 seconds). However, there was no significant difference between them ($t(59) = 1.983$, $p = 0.052$).

In order to analyze the task completion time in detail, we classify the time into two categories. They are a browsing time and an operation time. The browsing time is time to see the document or the figure with no operation. If there is an operation, such as scrolling or zooming, it is considered as the operation time.

With this classification, it is found that, as shown in Fig. 3, the operation time with the proposed method was 5.5 seconds and was significantly shorter than that with the conventional method (10.8 seconds, $t(59) = 4.6709$, $p < 0.01$).

Talking about the browsing time, there was no significant difference between two methods. In our prototype system of the proposed method, as mentioned before, the resolution of diagram area was less than that of document area. Especially, the height of diagram area was less than a half of the document height. With this prototype system, the size of pulled-in figure was adjusted to that of the floating window so that it was sometimes smaller than the size of figure with the conventional method. In such case, it might take a little bit long time to browse the figure. This might be one reason why there was no significant difference about the task completion time between two methods.

These results mean, anyway, that the proposed method is more efficient than conventional method for operation on mobile devices.

4.2 Subjective Evaluation

Fig. 4 shows the results of the subjective evaluation about ease of learning and ease of operation. From this figure, it is found that both methods were easy to learn and to operate. This means that the participants viewed the proposed method very positively.

It might be easy for participants to learn the operation of conventional method because the number of operation was so few, they were zoom in, zoom out, and scroll. Furthermore, participants were already familiar with such operation. Talking about the proposed method, it was also easy to learn the operation. This might be also caused by a few operations to learn. A few participants also commented that the pulled-in manner of the proposed method was very intuitive and easy to learn.

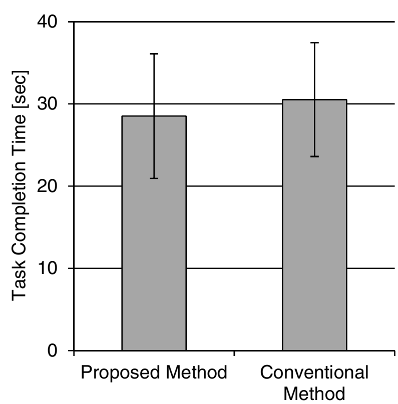


Fig. 2. Task completion time

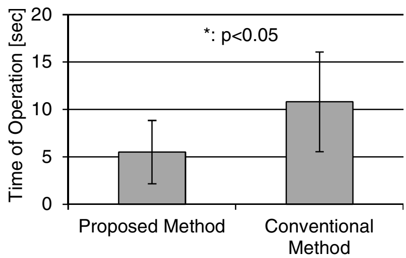


Fig. 3. Time of operation

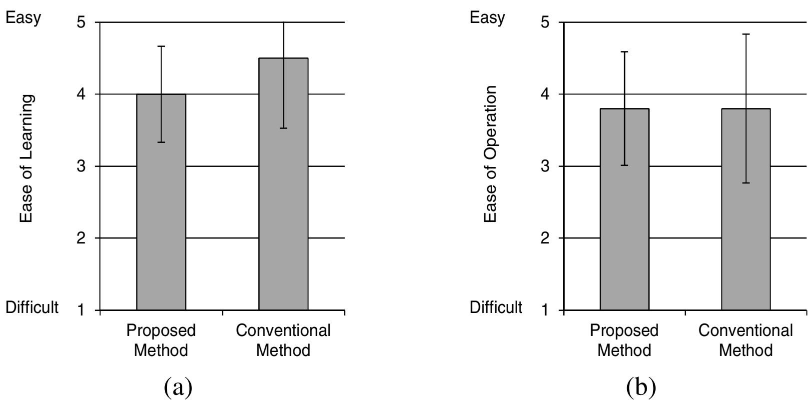


Fig. 4. Subjective evaluation. (a) Ease of learning. (b) Ease of Operation

The manner of operation with the proposed method was similar to that of conventional method. This might be a reason why there was no difference between the proposed method and conventional method about the ease of operation. Furthermore, some participants highly evaluated the proposed method because they could decrease the number of scrolling to see the figure and the document mutually.

Some participants complained that the border width of the floating diagram window was too narrow that they sometimes could not touch them as they wanted. They claimed to widen the width. However, to widen it makes the size of diagram smaller and might increase the difficulty of browsing the diagram.

5 Related Work

In order to improve the navigation on small screens, Igarashi et al. propose the speed-dependent automatic zooming. In this zooming, the view automatically zooms out when the user scrolls rapidly so that the perceptual scrolling speed in screen space remains constant [2]. Ishak et al. propose content-aware scrolling (CAS), an approach that takes into account various characteristics of document content to determine scrolling direction, speed, and zoom [3]. Both methods can improve the scrolling speed but they don't focus on browsing both referrer sentence and its referencing diagram mutually and efficiently.

Web pages are typically designed with the desktop screen in mind and can be hard to read on small screens. There are many studies to overcome this problem with adapting the original web page to the small screen. Summary Thumbnails [4] provides an enhanced thumbnails view with text fragment. Ahmadi et al. propose a method that automatically adapts a desktop presentation to a mobile presentation with generating a set of small subpages [1]. However, with these kind of adaptation, it should be hard to read referrer sentence and its referencing diagram mutually.

6 Conclusion

In this paper, an efficient method to browse the document including diagrams on mobile device is proposed and evaluated experimentally. From the experiment, it is found that the proposed method is efficient for operation on mobile devices. From the subjective evaluation, it is also found that the proposed method is viewed very positively.

Acknowledgments. This research was supported by a grant-in-aid for scientific research (C) (21500121, 2009-2011) from Japan Society for the Promotion of Science (JSPS).

References

1. Ahmadi, H., Kong, J.: Efficient web browsing on small screens. In: Proceedings of the Working Conference on Advanced Visual Interfaces, pp. 23–30. ACM, New York (2008)
2. Igarashi, T., Hinckley, K.: Speed-dependent automatic zooming for browsing large documents. In: Proceedings of the 13th Annual ACM Symposium on User Interface Software and Technology, pp. 139–148. ACM, New York (2000)
3. Ishak, E.W., Feiner, S.K.: Content-aware scrolling. In: Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology, pp. 155–158. ACM, New York (2006)
4. Lam, H., Baudisch, P.: Summary thumbnails: readable overviews for small screen web browsers. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 681–690. ACM, New York (2005)

Mobile Reminder for Flexible and Safe Medication Schedule for Home Users

Pei-Hsuan Tsai[1], Chi-Sheng Shih[2], and Jane W.-S. Liu[1]

[1] Inistiute of Information Science, Acadeima Sinica, Taipei, Taiwan
[2] Graduate Institute of Networking and Multimedia,
National Taiwan University, 106 Taipei, Taiwan
{peipei,janeliu}@iis.sinica.edu.tw, cshih@csie.ntu.edu.tw

Abstract. iMAT is a system of automatic medication dispenser and software tools. It is for people who take medications on long term basis at home to stay well and independent. The system helps its users to improve rigor in compliance by preventing misunderstanding of medication directions and making medication schedules more tolerant to tardiness and negligence. Medication schedule manager in iMAT can be deployed either on standalone automatic, intelligent medication dispensers, or on computers and smart phones accompanied with traditional pillboxes. In this paper, we present the design and implementation of PROMUS and the results of a user experience study. PROMUS is a medication scheduler manager and reminder for modern smart phone devices. Similar to the scheduler for the iMAT dispenser, PROMUS scheduler takes into account both user preferences and medication directions in generation of flexible medication schedules and compliance enforcement. In particular, we focus on the interactions between the user and medication schedule manager to avoid medication errors. PROMUS schedules medication events to be more flexible and friendly by grouping the medication doses to reduce the number of times medications are to be taken and allow the user longer response time. PROMUS also provides intuitive, visually appealing user interface, clear organization, and familiar terminology that can be acted upon in conformity with the original behavior of medication record keeping.

Keywords: Medication Scheduler, Smart Devices, and Medication Compliance.

1 Introduction

This paper describes a system of smart medication dispensers, medication schedule managers, and its implementation on smart phones. The system, called iMAT (intelligent medication administration tools), targets as users the growing population of elderly individuals and people with chronic conditions who are well enough to maintain active, independent lifestyles. Such a person may take many prescriptions and over the counter (OTC) medications and health supplements at home and work without close professional supervision. In subsequent discussions, by a user, we mean such a person.

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 107–116, 2011.

Nowadays, modern drugs can help people conquer previously fatal diseases, control debilitating conditions, and maintain wellness and independence well into old age, provided that the drugs are taken as directed. Unfortunately, even critically important drugs such as those for treatments of hypertension, diabetes, and heart conditions are often not taken as directed [1]. The fact is that statistics on health care quality continue to show alarmingly rates and serious consequences of preventable medication errors [2–5]. Administration errors due to non-compliance to medication directions are known to contribute 25-40% of all preventable medication errors and are the cause of approximately 10% of hospital admissions and 23% of nursing home admissions. The primary goal of iMAT is to prevent administration errors as much as possible and when errors occur despite prevention efforts, reduce the adverse effects caused by them. iMAT can also help to make sure that interactions among all medications of each user and their interactions with food and drink have been properly accounted for by the directions for the user.

A look at causes of non-compliance points out that information technology can help eliminate many common ones, including misunderstanding of medication directions, inability to adhere to complex medication regimens, and inconvenience of rigid schedules. iMAT is designed specifically to eliminate these causes. A user of iMAT medication dispenser and schedule manager has no need to understand the directions of her/his medications. iMAT enables the pharmacist of each user to extract a machine readable medication schedule specification (MSS) from the users prescriptions and OTC directions. Once loaded into an iMAT dispenser or schedule manager, the tool automatically generates a medication schedule that meets all the constraints specified by the users MSS. Based on the schedule, the tool reminds the user at the times when some doses should be taken and provides instructions on how the doses should be taken (e.g., with 8 oz of water, no food within 30 minutes, etc.) In this way, iMAT helps to make complex regimens easy to follow.

PROMUS is a variant of iMAT: It is designed as a medication scheduler manager and reminder for modern smart phone devices. Similar to the scheduler for iMAT medication dispenser [6], PROMUS scheduler takes into account both user preferences and medication directions in the generation of flexible medication schedules and enforcement of compliance. In particular, we focus on the interactions between the user and medication schedule manager to avoid non-compliance errors. PROMUS schedules medication events to be more flexible and user friendly by grouping the medication doses to reduce the number of times medications are to be taken and to allow the user longer response times to reminders. PROMUS also provides an intuitive, visually appealing user interface, clear organization scheme, familiar terminology that can be acted upon in conformity with user's original behavior of medication record keeping. To evaluate the effectiveness of PROMUS, we conducted a field trial on users with different backgrounds: different jobs, ages, genders and lifestyles. We logged all the user activities on the smart phone to understand how the users used the devices.

The remainder of the paper is organized as follows. In Section 2, we present an overview of iMAT and related works. Section 3 presents the design and implementation of PROMUS. Section 4 presents our evaluation results. Section 5 concludes the paper.

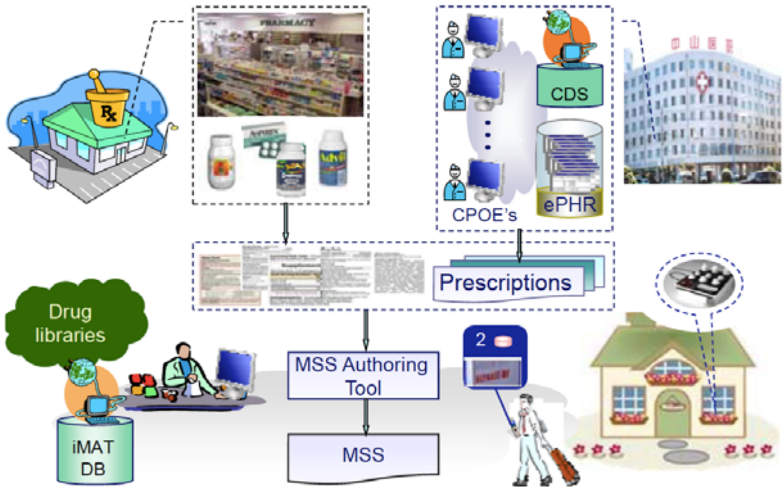


Fig. 1. iMAT in medication use tool chain

2 Overview of iMAT and Related Works

iMAT, as a distributed service, is designed specifically to eliminate the causes for medication non-compliance. A user of iMAT medication dispenser and schedule manager has no need to understand the directions of her/his medications so as to correctly take medications. Figure 1 shows how iMAT fits in a chain of information systems and tools for medication use process: It complements computerized physician order entry (CPOE) systems [7-12] at the top level by supporting the dispensing and administration stages of the process. By far, computerized physician order entry (CPOE) systems are the most well developed tools in the tool chain. Today, CPOE systems are used in a majority of hospitals and clinics in developed countries. Recent data on their effectiveness show that CPOE systems, together with clinical decision support (CDS) and electronic patient health and medication records (ePHR and eMAR) systems [13, 14], can help prevent up to 80% of prescription errors, i.e., 40% of all medication errors.

The MSS authoring tool and iMAT database, shown at the lower left half of Figure 1, are for pharmacists. An essential function of the authoring tool is to merge the directions of all medications of each user and generate from the merged direction a machine readable medication schedule specification (MSS) for the user. As stated earlier, the MSS is needed to guide the operations of user's medication dispenser and schedule manager. We describe the operations of a version of the tool in [15].

A user may have an iMAT medication dispenser for use at home, as shown in the bottom right corner of Figure 1. The medication scheduler that runs on the dispenser can serve as a schedule manager. It delivers reminders to a cell phone and other mobile devices, also shown in the figure. A user may choose to have only a schedule manager and have the tool run on a PC, laptop or a smart phone that can hold the MSS and has network access. These devices have the same purpose as numerous pillboxes and programmable medicine dispensers (e.g., [16-18]) for home use and mobile medication administration tools (e.g., [19, 20]) for use in hospitals and long-term care facilities. Existing pillboxes and dispensers require the user to load the individual doses of medications into the device, understand their directions and program the device to send reminders accordingly. This error-prone manual process and rigid medication schedules are serious disadvantages for users targeted by iMAT.

Intelligent medication advisory tools and services such as MEDICATE Tele-assistance System, Magic Medicine Cabinet, and other medication advice services [21-24] can check directions for drug interactions for users at home. Like schedules used by our dispenser and schedule manager, medication schedules used by these automatic devices and scheduling tools can also be adjusted to compensate for user tardiness and condition changes. The advices and adjustments are provided by care takers who monitor and supervise the user via Internet, however. Those devices are better suited for users who need close professional supervision and fully integrated health care services. In contrast, our medication dispenser and schedule manager are capable of making schedule adjustments permitted by existing prescriptions without requiring their users to incur the costs in fees and privacy loss of close monitoring and care.

3 Design and Implementation for Mobile Phone Reminder Applications

An advantage of the iMAT dispenser is that it can help, as much as an IT device can, to make sure that user retrieves the right dose of each medication from the right container when the user responds to reminder and comes to retrieve the medication. When the user is away from home, a portable medication reminder is required to reminder to take medications. Figure 2 shows two configurations of the portable schedule manager. The dotted box below the devices encircles the software components. Here, the user interface manager plays the role of the dispenser controller. It manages the interface to facilitate the interactions with the user according to the user's preference. Like a dispenser, the manager also maintains locally the user's current medication schedule specification and medication record.

PROMUS is a variant of iMAT: It is designed as a portable medication schedule manager that is deployed on smart phones. The current version is implemented in Java and is compatible with Android 2.1 or greater.

As discussed in previous reports, most, if not all, of the medication schedules highly depend on patient's daily schedule, such as the time to start a day and to take meals. When not necessary, PROMUS does not ask the users to change his/her daily schedule for medication compliance: It allows the users to provide their daily schedules so that the PROMUS scheduler can adjust each user's medication schedule to fit his/her daily schedule. Figure 3 shows the screenshot for configuring user preferences on PROMUS. The users provide their schedules for wake-up, breakfast, lunch, dinner, and go-to-bed. These parameters will be used as input parameters for medication scheduler. The users can also choose the preferred mode(s) such as vibration, rings, or music for reminders.

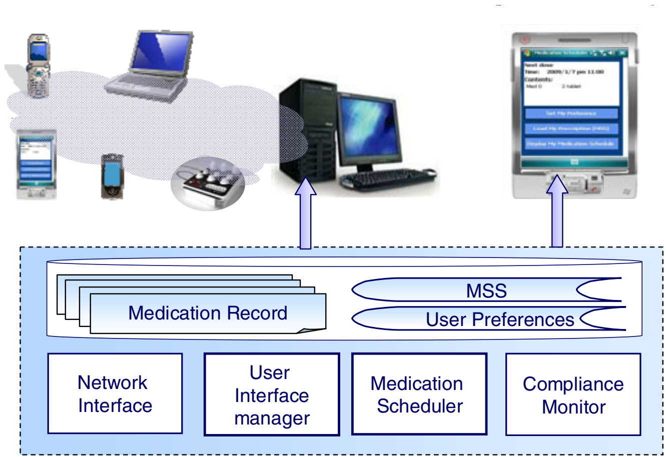


Fig. 2. Alternative Configuration of iMAT Medication Scheduler

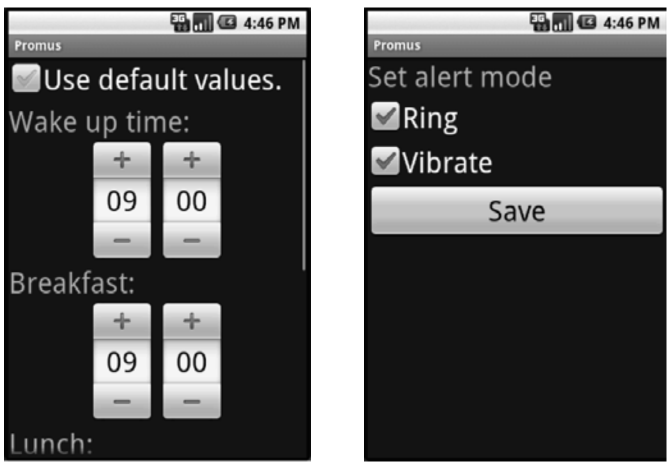


Fig. 3. User Preference Configuration

PROMUS periodically checks if there is any medication due to be taken. The period is now set to one hour and can be adjusted in the configuration file. One hour is most practical for the users not staying in health care institutes. When some medication is scheduled, PROMUS sends reminder(s) to the user. (PROMUS now provides three types of notifications: *pop-up notification window*, *vibration* and *rings*.) The user can select the preferred notification mode in preference configuration. In pop-up windows, there are three options to choose: “Take medications now”, “Remind me later”, and “Skip this dose”. If the user chooses “Take medications now”, PROMUS displays the directions including medication names, dosages, colored photos of the medications and special instructions for the medications. The information for the scheduled medications is displayed for one medication at a time by to avoid confusion. Figure 4 shows the screenshot for the medication information and medication prescriptions. The medication directions stored in PROMUS, including names of medications (Chinese name, English name and scientific name), granularity of medications, total amounts of medications, usage, dosage, warning, therapy and side effects are from the Health Insurance Bureau of Taiwan government. One can use different databases to provide such information. For each scheduled medication, the user can choose to take or to skip it. PROMUS records the user's choices in the user's medication record. If the user decides to take medication later, PROMUS suspends notifications for a user-defined postpone interval. The default postpone interval is 5 minutes and can be changed in preference configuration. If the user decides to skip the dose, PROMUS stops sending notifications and waits until next schedule point.

The scheduled time for a medication is only a candidate time instance to take the medication. Often, there are numerous feasible schedules conforming to the medication schedule specification, [25]. PROMUS takes advantage of this fact to provide the user with the option to check whether he/she can take medications before scheduled times. In this mode, the user can take medications within a tolerable interval before the scheduled time. On PROMUS, the user can choose "Available medications now". PROMUS checks and displays the medications that can be taken at that time. (Again, the user can choose to take it or to skip it by pressing the button on the screen, and PROMUS would record all the user's actions automatically.)

We use four frequency tables to represent the feasible medication schedule times. They are the number of days for a prescription period, number of days to take medications per period, number of dose times in a day and intervals between each doses. Take as an example, the direction "Take every next day, three times a day, after meals". This direction indicates that the period for the prescription is two days, the elapsed interval for each prescription is one day, the number of dose in a day is three, and the intervals between two consecutive doses are four to six hours. Hence, when the user takes meals at 8AM, 12PM and 5PM, the doses will be scheduled at 9AM, 1PM and 6PM, respectively.

Fig. 4. Prescription and Medication Information Review

The PROMUS scheduling algorithm uses a data structure to represent the schedule and status for each medication: The structure contains “Schedule”, “Allowable time slots”, “Record”, and “Status”. “Schedule” is generated based on the dosages specified by the medication directions, frequency table and user preference. Usually, feasible time to take medications is not a single point of time. Instead, it is a time interval, which is represented by a number of time slots. In addition, most medication directions include instructions such as “If you miss the dose, please take it as soon as possible. If the time is close to next dose, please skip this one and go back to your regular schedule. Do not take it twice.” Such instructions give PROMUS the flexibility in computing the time interval for user to take medications and the time intervals for not taking medications.

For the example described above, the allowable time slots for user to take medications would be 9:00 am to 12:00pm, 1:00pm to 4:00pm and 6:00pm to 2:00am. Figure 5 shows the data in PROMUS on this schedule.

Time (hour)	0	1								9				13					18					
Schedule	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0
Allowable time slots	1	1	0	0	0	0	0	0	0	1	1	0	0	1	1	1	0	0	1	1	1	1	1	1
Initial state																								
Record	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Status	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
After user takes medication																								
Record	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Status	1	1	1	1	1	1	1	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1

Fig. 5. An example of schedule tables

“Record” is used to record the time that the user takes the medication. “Status” is used to avoid overdoses and is set to false after user takes medications. Every time when the user responds that he/she takes the medication, PROMUS records the action and updates the status from available to unavailable of the medication by changing the element values of “Status” array to zero. The time granularity of the time tables are set to one hour, which is consistent with the period for the timer to check the prescriptions.

4 Field Trials and Evaluation Results

To evaluate the effectiveness of PROMUS and to build incrementally a record of use patterns, we select ten users with different education background, job types, and ages. Their ages range from 25 to 60 years and their occupations include students, engineers, housewives, retirees, and businessmen. The medications they took include those for diabetes, hypertension, heart disease, asthma, and other chronic diseases. In the experiment, both the prescriptions issued by physicians and extracted from OTC directions are stored in PROMUS. Each subject uses PROMUS for five to seven days.

Before the subject starts to use PROMUS, we logged their medication compliance using questionnaires on paper. According to the records, on average, they take 5 to 8 kinds of medications and take medications 3 to 5 times a day. The records also show that their average missing dose rate is approximately 1/3 for prescription medications and the average delay time in take medications is approximately 1 to 3 hours. Due to the space limit, we describe below three cases in the experiment.

Case 1: The subject is a 30-year-old male who is an office staff in a science and technology company and uses smart phones daily. He is a diabetic and takes 11 types of medications. He takes medications at least four times a day. Although his daily schedule is very regular, he often forgets to take medications in the morning. Sometimes, when he remembered to take medications later, he took the medications without paying attention to see whether the time is close to the next dose time. Consequently, he either missed the dose or took overdose in those occasions. To maintain the stability of blood sugar for a diabetic is very important. Overdose may lead to low blood sugar levels and cause the patient faint. According to the user's response, PROMUS reduces the chances of missing medications in the morning and helps him not to overdose. He found that the user interface is clean and the operation is simple without error-prone human input.

Case 2: The subject is a 24-year-old female and is a graduate student with Computer Science major. She is an asthma patient and never uses smart phones before. She has two medications to take and must take them every 12 hours for controlling her asthma. She also takes vitamins in an irregular schedule. However, her daily schedule is not as regular as the others. She sometimes stays up all night and sleeps for more than 12 hours in the next day. As a result, she misses medications from time to time. According to her logged activities, she missed some doses and later took medications when she did not feeling well or until she got the asthma attack. Asthma attack is very dangerous and acute asthma may cause death. Having a regular medication schedule is important for them to control the asthma and decrease the attack probability. PROMUS identified the medication for controlling asthma is critical and forced her to wake up take the medication even when she was sleeping. According to her feedback, PROMUS repeatedly sent reminder until the medications are take and helped her to take medication at right time. The chances for missing medication were significantly lowered.

Case 3: The subject is a 60-year-old businessman and has both hypertension and diabetes. He has eight types of medications for his diseases and usually takes medications three times a day. He never uses smart phones and is not interested in learning how to use new IT devices. Due to his tight schedule and frequent business meetings, even when he remembers to take medications, he still misses doses because it is inconvenient for him to take medications in front of his colleagues and visitors. Later he would either skip medications or take all the medications including missed medications and medications scheduled on the next dose time. His feedbacks show that the best part of PROMUS is that he does not need to spend much time to learn how to use it. After having PROMUS, he relies it to follow the medication directions. PROMUS reminds him and shows the medications he can take later without asking him to reschedule the dose time. In this way, PROMUS decreases the miss dose rate and helps him to take medications correctly. The operation flow requires the least user input, and the screen shows the concise and necessary information, which save his time to understand and then responds.

The experimental results show that PROMUS does improve medication compliance for those users who are busy and may forget to take medications, and users who are not sure when should take medications or what should be taken. However, for those users who do not want to take medications because of other mental and physical factors. For example, one may hate to swallow medications; the other one do not think he/she needs to take medications. These users may deliberately ignore the notifications for taking medications sent by smart phone or intentionally fraud by responding that they have taken medications. PROMUS cannot force them to take medications and is unaware of fraud. Modern smart phones often run out of battery, and the devices cannot perform any function when they are in the off state. In this situation, PROMUS cannot work and needs the user to provide his/her medication record when it can be turn on and working again. There is a higher risk that user may miss doses when PROMUS does not work since they already have the habit to rely on PROMUS to remind them to take medications. In addition, some medications are prescribed to take when necessary instead of being taken regularly. These medications usually require the user to measure some vital sign to determine whether to take or not the medication. PROMUS cannot provide any assistant on this matter.

5 Conclusions and Future Works

PROMUS is designed as a medication scheduler manager and reminder for modern smart phone devices. As a portable implementation of iMAT, both user preferences and the medication directions specified by prescriptions are taken into account to carry out compliant and flexible medication schedules. In particular, we focus on the interactions between the user and medication schedule manager to avoid medication errors. The experimental results show that PROMUS does help to improve medication compliance for certain types of users but not all. The major factor is whether the users are comfortable for using smart phones as a reminding device. Few of our test subjects prefer to have vocal feedback and reminder, rather than text or graphic base reminders. Last, providing the advice and assistant for taking PRN medication will be added to PROMUS.

References

1. Office of Inspector General, U.D.O.H., Resources, H.: Medication regimens, causes of noncompliance (1990)
2. Food, U., Administration, D.: Drug safety and availability. Medication errors (2009)
3. Joint Commission of Accreditation Program. National patient safety goal on reconciling medication information (2010)
4. Veacez, P.: An individual based framework for a study on medical error. International Journal for Quality in Health Care 18 (2006)
5. Lisby, M., Nielsen, L.P., Mainz, J.: Errors in the medication process: frequency, type, and potential clinical consequences. International Journal for Quality in Health Care 17 (2005)
6. Tsai, P.H., Yu, C.Y., Wang, W.Y., Zao, J.K., Yeh, H.C., Shih, C.-S., Liu, J.W.S.: iMAT: Intelligent Medication Administration Tools. In: The Proceedings of IEEE Healthcom (July 2010)

7. CPOE.org, http://www.cpoe.org/
8. Kuperman, G.J., Bobb, A., Payne, T.H., et al.: Medication related clinical decision support in computerized provider order entry systems: A Review. J. Am. Med. Inform. Assoc. (2007)
9. Cutler, D.M., Feldman, N.E., Horwitz, J.R.: U. S. Adoption of Computerized Physician Order Entry Systems. Health Affairs 24(6) (2005)
10. Davis, R.L.: Computerized physician order entry systems: the coming of age for outpatient medicine. PLoS Medicine (September 2005)
11. Koppel, B., et al.: Role of computerized physician order entry systems in facilitating medication errors. Journal of AMA 293(10) (2005)
12. King, W.J., Paice, N., et al.: The effect of computerized physician order entry on medication errors and adverse drug events in pediatric inpatients. Pediatrics (September 2003)
13. Electronic Medication Administration Record (eMAR) (2009), http://www.dh.org/body.cfm?id=434&oTopID=434
14. Baron, R.J., Fabens, J.E.L., Schiffman, M., Wolf, E.: Electronic health records: just around the corner? Or over the cliff? Annuals of Internal Medicine (2005) Health Information Systems, http://www.hhs.gov/healthit/ahic.html
15. Yeh, H.C., Hsiu, P.C., Shih, C.S., Tsai, P.H., Liu, J.W.S.: APAMAT: a prescription algebra for medication authoring tool. In: Proceedings of IEEE International Conference on Systems, Man and Cybernetics (October 2006)
16. e-pill, Pill Dispenser, http://www.epill.com/dispenser.html and MD 2, http://www.epill.com/md2.html
17. Pill boxes, http://www.dynamicliving.com/automated_medication_dispenser.htm
18. My Pill Box, http://www.mypillbox.org/mypillbox.php
19. Rx Showcase, http://www.rxinsider.com/prescription_dispensing_automation.htm
20. Tsai, P.H., Chuang, Y.T., Chou, T.S., Liu, J.W.S., Shih, C.S.: iNuC: an intelligent mobile nursing cart. In: the Second International Conference on BioMedical Engineering and Informatics (October 2009)
21. Nugent, C., et al.: The next generation of mobile medication management solutions. International Journal of Electronic Heathcare 3(1) (2007)
22. Governo, M., Riva, V., Fiorini, P., Nugent, C.: MEDICATE teleassistance system. In: The 11th International Conference on Advance Robotics (June 2003)
23. Murray, M.D.: Making health care safer: a critical analysis of patient safety, 01-E58, Agent for Healthcare Research and Quality. In: Automated Medication Dispensing Devices (2001)
24. Wan, D.: Magic Medicine Cabinet: A Situated Portal for Consumer Healthcare. In: Gellersen, H.-W. (ed.) HUC 1999. LNCS, vol. 1707, p. 352. Springer, Heidelberg (1999)
25. Tsai, P.-H., Yeh, H.C., Yu, C.Y., Hsiu, P.C., Shih, C.-S., Liu, J.W.S.: Compliance Enforcement of Temporal and Dosage Constraints. In: Proceedings of IEEE Real-Time Systems Symposium (December 2006)

Enabling Efficient Browsing and Manipulation of Web Tables on Smartphone

Wenchang Xu and Yuanchun Shi

Department of Computer Science and Technology, Tsinghua University,
Beijing, P.R. China
wencxu@gmail.com, shiyc@tsinghua.edu.cn

Abstract. Tables are very important carriers of the vast information on the Internet and are widely used in web pages. However, most designs of web tables are only for desktop PCs and just focus on how to visually and logically show large amount of data without considering their visual effects on small-screen devices. Therefore, users suffer inconvenience when browsing web tables on smartphone. In this paper, we propose to enable efficient browsing and manipulation of web tables on smartphone in order to solve the problems of both information retrieval and content replication from web tables. We implemented a mobile web browser on Android 2.1 platform, which deals with web tables in three steps: genuine table detection, table understanding and user interface design. We conducted a user study to test the effects that users used such tool. Experimental results show that the tool increases users' browsing efficiency of web tables and the novel browsing and manipulation modes are well accepted by users.

Keywords: Web table, Table extraction, Table understanding, Table manipulation, User interface design.

1 Introduction

Tables are very important carriers of the vast information on the Internet and are widely used in web pages. As the rapid development of the smartphone, more and more people begin to surf the Internet with a mobile web browser on smartphone. However, most designs of web tables are only for desktop PCs and just focus on how to visually and logically show large amount of data without considering their visual effects on small-screen devices. Therefore, users suffer inconvenience when browsing web tables on smartphone.

Generally speaking, the limitation of browsing web tables on smartphone with small screens mainly reflects in the following two aspects:

- The Information retrieval from web tables is inconvenient. When users want to find certain information in a table or to compare data in different columns or rows, they have to scroll the page repeatedly in order to make sure of the column or row header and confirm the target, which is a waste of time, especially when the table is large or the compared columns or rows are distant from each other.

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 117–126, 2011.

- The content maintenance and replication of web tables are time-consuming. In this regard, when users want to copy certain part of a web table into a local file, they usually have to copy the cell content one by one, which may take a long time and usually cause errors. There is no express way to save the web table quickly as a local file of table format such as Excel that is easy for further editing.

Existing research [12][15] has been devoted to addressing the problem, which proposes to enable interactive access to web tables. However, they either lack practical solutions for mobile devices [15], or ignore the second limitation and fail to consider the popular finger-touch interactive mode of existing smartphone during the interaction design [12].

In this paper, we propose to enable efficient browsing and manipulation of web tables on smartphone in order to solve the problems of both information retrieval and content replication from web tables. We choose the Android 2.1 platform and design a mobile web browser based on the WebView object. On this basis, we implement all logical functions by JavaScript, which will be added to a web page when it is loaded in our web browser. Our implementation includes three modules, which are genuine table detection, table understanding and user interface design. A genuine table is used to display logically related data with significant semantics, as is defined in previous work [13]. We detect genuine tables from web pages based on machine learning which extracts 7 layout features and 8 content features as well as the specific word group feature. We did data training in advance and wrote the training result into a JS file so that the classification can be executed in real time. We then classify genuine web tables into three sub-categories according to their header styles, including column-wise, column-row-wise and row-wise. We identify each cell of the tables as one of the 7 basic data types, which are image, form, hyperlink, alphabetical, digit, empty and others. With the table understanding results, we design the user interface of the tool. According to the requirements, we design both browsing mode and manipulation mode, each of which contains 5 functions that facilitate users' browsing experience of web tables on touchscreen-based smartphone. Specifically, the functions of the browsing mode are executed in tables of original web pages while the functions of the manipulation mode are executed in generated tables on new application panels.

In addition, we conducted a user study to test the effects that users used such tool. We recorded the frequency they used each function and the time they finished given tasks with or without the tool. We also designed a questionnaire to collect users' evaluations and use experience. Experimental results show that the tool increases users' browsing efficiency and the novel browsing and manipulation modes are well accepted by users.

The rest of the paper is organized as follows. First, we review existing work related to both our motivation and implementation. We then present the implementation of the tool, introducing both the analytical approaches and the user interface design. Next, we describe the user study and the experimental results. We finally discuss about future work and draw conclusion.

2 Related Work

We review existing research on both web table analysis and interactive user interfaces of accessing web tables that are related to our work.

2.1 Web Table Analysis

Web table analysis has attracted much attention from researchers in the areas of Web Data Mining and Information Retrieval since late 1990s [2][3][5][6][7][13][16][17]. In recent years, some researchers even try to extend table analyzing technologies to other document formats like PDF files [4][8][9][10]. Lim and Ng [7] proposed to automatically retrieve hierarchical data from HTML tables by constructing the content tree for each table, without pre-requiring the internal table structure. Yoshida et al. [17] presented a new approach to integrate web tables describing similar objects based on table structure recognition. Cohen et al. [2] developed an extensible wrapper-learning system called WL2 to exploit HTML tables and lists. Yang and Luk [16] first presented the definition of Web Table Mining and developed a frame work for comprehensively analyzing the structural aspects of web tables. Hurst [5] noticed that HTML TABLE elements were more and more often used to control the layout of web pages and tried to identify true web tables by learning from HTML DOM and geometric features. Wang and Hu [13] focused on web table detection and proposed to automatically classify web tables either as genuine or non-genuine by machine learning. Krüpl and Herzog [6] also concentrated on detecting genuine web tables, but they relied on the visual rendition of web pages rather than the HTML code. Later, Gatterbauer et al. [3] extended the idea of visually guided web table detection and used a model of the visual representation of web pages to extract domain-independent information from web tables.

Our target of web table analysis is to detect genuine web tables and understand their structure and content automatically. Due to the limited computation capabilities of mobile devices, we adopt the simple but efficient machine learning based approach proposed by Wang and Hu [13] to complete the genuine table detection.

2.2 Interactive Access to Web Tables

To improve the efficiency of web table browsing, more and more attention is paid to the interactive browsing and manipulation of web tables.

Asakawa and Itoh [1] developed a non-visual web table navigation method enabling both horizontal and vertical navigation with a table cursor, a table pointer and a cell-jumping key. However, they only dealt with gridded tables, that is, TABLE elements defined in HTML documents but without COLSPAN and ROWSPAN. Tajima and Ohnishi [12] proposed three modes for web table browsing on small screens: normal mode, record mode and cell mode. They concentrated on how to present a segment of a large web table as the user requires, without concerning about the relationships among data in different table cells, and they did not present any user evaluation. Yang and her colleagues proposed to enable interactive access to web tables. They implemented a plug-in for Microsoft Internet Explorer that provided a user interface for the functionality [14][15]. An extension for Firefox named TableTools [11] was also developed to provide similar functionality. However, there is still no interactive browsing and manipulation tool of web tables actually deployed on mobile devices.

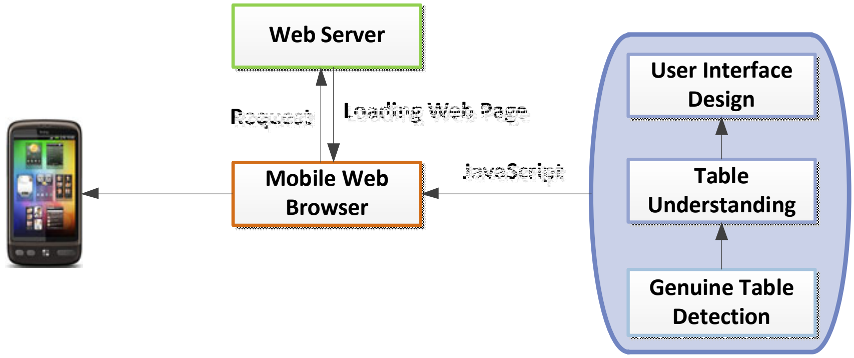


Fig. 1. System structure of the implementation

3 Implementation

To address the problems of both information retrieval and content replication from web tables on small touchscreen based mobile devices, we propose to enable efficient browsing and manipulation of web tables on smartphone. Figure 1 shows the structure of our implementation. We implement a mobile web browser based on the WebView object on the Android 2.1 platform. When a user enters a URL in the mobile web browser, the web browser sends a request to the target web server. It then loads the web page returned from the web server and renders various web page elements. After the page is loaded, a piece of JavaScript code will be executed automatically, which detects genuine tables in the web page, understands table structures and provides the user interface, through which the user can browse and manipulate web tables interactively. We will next introduce the three central steps, genuine table detection, table understanding and user interface design in detail respectively.

3.1 Genuine Table Detection

Genuine table detection is the first step of the implementation. We denote each web table as either genuine or non-genuine according to previous work [3][6][13]. Generally speaking, non-genuine tables are table elements used for designing page structure layout, while genuine tables are table elements used for displaying data in the two-dimensional manifestation. Here we classify detected tables into either genuine or non-genuine by machine learning. We partially adopt the automatic web table detection method proposed by Wang and Hu [13] which summarizes 15 features of web tables, including 7 layout features and 8 content features. Layout features are calculated based on row numbers, column numbers and cell length, while content features are calculated based on the number of different content types of table cells, such as image, form, hyperlink, alphabetical and so on. In addition, we improve and add the simplified word group feature, which describes detailed document content of web tables.

We compare four different classification algorithms including Navie Bayesian, Decision Tree, SVM with linear kernel and SVM with RBF kernel with the above extracted features. The training data set is constructed from 1774 web tables. By manual classification, we find that genuine tables account for only about 13.13% (233/1774) of all web tables, which is similar to conclusions in [13]. We define Precision as the proportion of genuine tables in all tables that are detected as genuine, and Recall as the proportion of tables which are detected as genuine in all genuine tables. Table 1 show the Precision and Recall values of different classification algorithms.

Table 1. Values of precision and recall of different classification algorithms

Algorithm	Precision	Recall
Navie Bayesian	91.80%	94.79%
Decision Tree	97.50%	94.25%
SVM(linear)	91.39%	93.91%
SVM(RBF)	95.81%	95.98%

Experimental results show that SVM with RBF kernel performs the best, with precision and recall achieving as high as 95.81%, 95.98% respectively. However, due to the limited computational capabilities of mobile devices, we use the classification algorithm of Navie Bayesian, which is more simple and efficient. Compared to our previous work [14], the precision and recall are increased by 1.3% and 0.8% respectively after the word group feature is added. We did the training in advance and wrote the training result into a JS file so that the classification can be executed in real time.

3.2 Table Understanding

After a genuine table has been detected, we will further carry out the table understanding in the following three aspects: determining table type, identifying cell data type and storing table content with our own data structure.

For the first aspect, we classify genuine web tables into three-sub-categories according to different table header styles, including column-wise, column-row-wise and row-wise. The type of each genuine table is decided depending on the layout of TH elements within it. If a genuine table contains no TH element, it will be classified as default (column-wise) and each cell in the first row will be regarded as a table header.

For the second aspect, we define seven basic data types for table cells, including image, form, hyperlink, alphabetical, digit, empty and others, each of which corresponds to the content type of the same name used in the genuine table detection. The type of each cell in a genuine table is identified based on six heuristics, namely H1-H6, as is shown in Table 2.

For the third aspect, we denote each genuine table as a data matrix. The width of the matrix is equal to the maximum cell count in a row in the genuine table, while the height of the matrix is equal to the row count of the genuine table. Each element of the matrix represents a corresponding table cell in the genuine table. Table cells with ROWSPAN or COLSPAN tags are represented by multiple matrix elements of the same properties. Each matrix element is recorded as an attribute-value pair with the header information, denoted as <<data-type, content>, isHeader, headerInfo>. Compared to the DOM structure of a TABLE element, our customized structure can be traversed with much lower cost.

Table 2. Heuristics for identifying cell data type

No.	Description
H1	If the cell contains FORM element, the cell is identified as form, otherwise use H2.
H2	If the cell contains IMG element, the cell is identified as image, otherwise use H3.
H3	If the cell contains A element, the cell is identified as hyperlink, otherwise use H4.
H4	If the inner text length (not including blanks) of the cell is 0, the cell is identified as empty, otherwise use H5.
H5	If the inner text of the cell is consisted of only digital numbers, the cell is digit, otherwise use H6.
H6	If the inner text of the cell is consisted of only digital numbers and alphabetical characters, the cell is alphabetical, otherwise the cell is others.

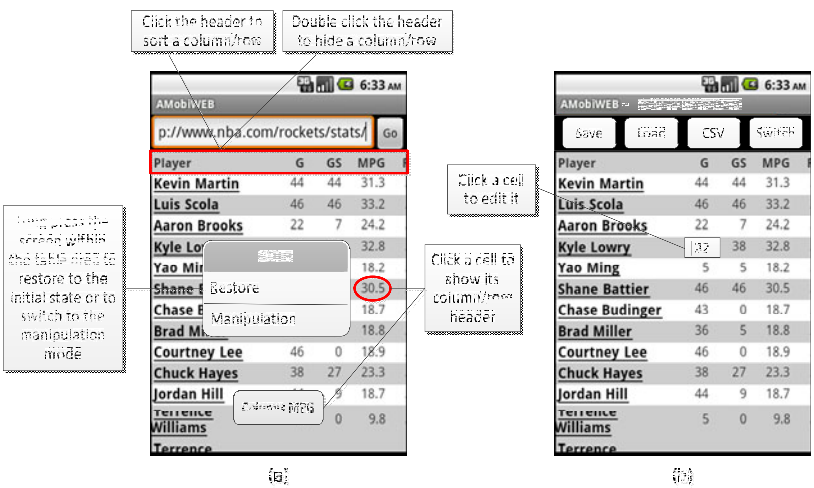


Fig. 2. User interface of the implementation: (a) browsing mode; (b) manipulation mode

3.3 User Interface Design

After genuine table detection and table understanding, we provide both browsing mode and manipulation mode according to users' requirements, each of which contains 5 functions that facilitate users' browsing experience of web tables on touch-screen-based smartphone. Specifically, the functions of the browsing mode are executed in tables of original web pages while the functions of the manipulation mode are executed in generated tables on new application panels. Figure 2 shows the user interface of our implementation and Table 3 lists all functions in both browsing mode and manipulation mode.

Table 3. Functions and their start methods of each mode

Mode	Function	Start Method
Browsing	**FB1**: Sort column/row	Click the header
Browsing	**FB2**: Hide column/row	Double click the header
Browsing	**FB3**: Show cell header	Click the cell
Browsing	**FB4**: Restore to the initial state	Select the menu item that is generated through long press within the table area
Browsing	**FB5**: Switch to the manipulation mode	Select the menu item that is generated through long press within the table area
Manipulation	**FM1**: Edit table cell	Click the cell
Manipulation	**FM2**: Save the table in the application	Click the button
Manipulation	**FM3**: Select a saved table in the application	Click the button
Manipulation	**FM4**: Save the table as a local CSV file	Click the button
Manipulation	**FM5**: Switch to the browsing mode	Click the button

4 User Study

We conducted a user study to test the effects that users used the tool and collected their evaluations to our design. We assume that the tool can greatly reduce users' efforts in typical cases as: 1) information retrieval from a large web table, e.g. finding a certain cell or comparing distant columns/rows; 2) content replication from a web table. Based on the assumption, we designed two scenarios and assigned specific tasks in the scenarios for browsing web tables, as is shown in Table 4.

Table 4. Scenarios and tasks designed for browsing web tables on smartphone

No.	Scenario	Task
1	You are browsing a 10-column genuine web table which displays the 20 teams of Spanish Football League ordered by their rankings (scores), with team names showed in col. 2, goals showed in col. 7 and scores showed in col. 10.	Find the teams with the largest and smallest number of goals; Find the teams with a score between 20 and 30.
2	You are browsing a 5-column genuine web table which displays the top 20 popular songs ordered by their rankings.	Save the table as a local CSV file.

Table 5. Results of the user study

Task	Tool	Average Times of using FB1-FB5 and FM1-FM5										Average Time Spent
		FB1	FB2	FB3	FB4	FB5	FM1	FM2	FM3	FM4	FM5	
1	On	1.8	7.2	5.8	0.2	0.0	0.0	0.0	0.0	0.0	0.0	21.542s
	Off	-	-	-	-	-	-	-	-	-	-	48.243s
2	On	0.0	0.0	0.0	0.0	1.0	0.0	0.8	0.2	1.3	0.3	14.131s
	Off	-	-	-	-	-	-	-	-	-	-	101.413s

Table 6. Results of the questionnaire. Helpful: 1 denotes not helpful at all and 5 denotes extremely helpful; Easy to use: 1 denotes difficult to use and 5 denotes extremely easy to use; Willingness to use: 1 denotes not willing to use at all and 5 denotes extremely willing to use.

Evaluation	FB1	FB2	FB3	FB4	FB5	FM1	FM2	FM3	FM4	FM5
Helpful	4.0	3.7	3.5	3.3	3.3	3.2	3.5	3.7	4.0	3.2
Easy to use	4.3	3.8	3.7	3.0	3.2	3.5	3.7	3.8	3.8	3.3
Willingness to use	4.2	3.7	3.3	3.2	3.5	3.2	3.7	3.5	4.2	3.2

We recruited 6 participants, all of whom were graduate students familiar with web browsing on mobile devices. They were showed a demo on how to use the tool for both browsing mode and manipulation mode in advance and were then required to complete each task in two ways: 1) using the default web browser of Android; 2) using our mobile web browser with the tool enabled.

We recorded the frequency that participants used each function and the time they finished given tasks with or without the tool. Table 5 shows the results calculated from automatically recorded timestamps of relevant operations for browsing and manipulating web tables. FB1-FB5 and FM1-FM5 are defined in Table 3.

From the results, we can conclude that FB1 (sorting column/row), FB2 (hiding column/row) and FB3 (showing cell header) are frequently used in Task 1 and FM4 (saving the table as a local CSV file) provides a shortcut for Task 2. The browsing mode and the manipulation mode can greatly decrease users' task completion time of Task 1 and Task 2 respectively.

We also designed a questionnaire to collect users' evaluations and use experience. We asked the 6 participants to rate the helpfulness and convenience levels as well as their willingness of using each function with a score between 1 and 5. Table 6 shows users' average results of the questionnaire. FB1 (sorting column/row), FB2 (hiding column/row), FB3 (showing cell header) and FM4 (saving the table as a local CSV file) get high scores, which is corresponding to the results of the user study. FM2 (saving a table in the application) and FM3 (selecting a saved table in the application) are not involved in the user study, but still get high scores.

Therefore, according to the results of both the user study and the questionnaire, we can conclude that the tool increases users' browsing efficiency and the novel browsing and manipulation modes are well accepted by users.

5 Conclusion and Future Work

In this paper, we propose to improve users' browsing experience of web tables on smartphone by presenting efficient analytical approaches and novel user interface that enable interactive browsing and manipulation of web tables. We implemented a mobile web browser on Android 2.1 platform, which deals with web tables in three steps: genuine table detection, table understanding and user interface design after a web page is loaded. We provide both browsing mode and manipulation mode to satisfy users' requirements of both information retrieval and content replication from web tables. We conducted a user study to test the effects that users used such tool. Experimental results show that the tool increases users' browsing efficiency of web tables and the novel browsing and manipulation modes are well accepted by users.

In the future, we are aiming to improve the user interface based on suggestions from the user study. Also we would like to extend the functions in order to provide users with richer browsing experience of web tables.

Acknowledgments. This paper was supported by MOE High-Tech Research and Development Plan of China under Grant No. 2009AA01Z336.

References

1. Asakawa, C., Itoh, T.: User Interface of a Non-visual Table Navigation Method. In: ACM SIGCHI Conference on Human Factors in Computing Systems (CHI 1999), pp. 214–215 (1999)
2. Cohen, W.W., Hurst, M., Jensen, L.S.: A Flexible Learning System for Wrapping Tables and Lists in HTML Documents. In: 11th International Conference on World Wide Web (WWW 2002), pp. 232–241 (2002)
3. Gaterbauer, W., Bohunsky, P., Herzog, M., Krüpl, B., Pollak, B.: Towards Domain-Independent Information Extraction from Web Tables. In: 16th International Conference on World Wide Web (WWW 2007), pp. 71–80 (2007)
4. Hassan, T., Baumgartner, R.: Table Recognition and Understanding from PDF Files. In: 9th International Conference on Document Analysis and Recognition (ICDAR 2007), pp. 1143–1147 (2007)
5. Hurst, M.: Classifying TABLE Elements in HTML. In: 11th International Conference on World Wide Web (WWW 2002), Poster Paper (2002)
6. Krüpl, B., Herzog, M.: Visually Guided Bottom-Up Table Detection and Segmentation in Web Documents. In: 15th International Conference on World Wide Web (WWW 2006), pp. 933–934 (2006)
7. Lim, S.J., Ng, Y.K.: An Automated Approach for Retrieving Hierarchical Data from HTML Tables. In: 8th ACM International Conference on Information and Knowledge Management (CIKM 1999), pp. 466–474 (1999)

8. Liu, Y., Bai, K., Mitra, P., Giles, C.L.: Automatic Searching of Tables in Digital Libraries. In: 16th International Conference on World Wide Web (WWW 2007), pp. 1135–1136 (2007)
9. Pinto, D., McCallum, A., Wei, X., Croft, W.B.: Table Extraction Using Conditional Random Fields. In: 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2003), pp. 235–242 (2003)
10. Ramel, J.Y., Crucianu, M., Vincent, N., Faure, C.: Detection, Extraction and Representation of Tables. In: 7th International Conference on Document Analysis and Recognition (ICDAR 2003), pp. 374–378 (2003)
11. TableTools, `https://addons.mozilla.org/zh-CN/firefox/addon/2637?src=reco`
12. Tajima, K., Ohnishi, K.: Browsing Large HTML Tables on Small Screens. In: Proceedings of UIST, pp. 259–268 (2008)
13. Wang, Y.L., Hu, J.Y.: A Machine Learning Based Approach for Table Detection on the Web. In: Proceedings of the 11th International Conference on World Wide Web (WWW 2002), pp. 242–250 (2002)
14. Xu, W., Yang, X., Shi, Y.: A New Mode of Browsing Web Tables on Small Screens. In: Proceedings of UCS, pp. 75–79 (2009)
15. Yang, X., Xu, W., Shi, Y.: Enabling Interactive Access to Web Tables. In: HCII 2009, pp. 760–768 (2009)
16. Yang, Y.C., Luk, W.S.: A Framework for Web Table Mining. In: 4th ACM CIKM International Workshop on Web Information and Data Management (WIDM 2002), pp. 36–42 (2002)
17. Yoshida, M., Torisawa, K., Tsujii, J.: A Method to Integrate Tables of the World Wide Web. In: 1st International Workshop on Web Document Analysis (WDA 2001), pp. 31–34 (2001)

Part II
Interaction in Intelligent Environments

User Interface Framework for Ambient Intelligence Platforms

Patricia Abril-Jiménez[1], Cecilia Vera-Muñoz[1],
María Teresa Arredondo Waldmeyer[1], Haydee Alvarez[2],
and José R. Baragaño Galán[3]

[1] Life and Supporting Technologies, Universidad Politécnica de Madrid
Avenida Complutense nº 30, Ciudad Universitaria, 28040 – Madrid, Spain
[2] SIEMENS, S.A.
División MED-HS, Parque Empresarial La Carpetania.
Avenida Leonardo Da Vinci nº 15-C1ª Planta, 28906 Getafe – Madrid, Spain
[3] Universidad Politécnica de Madrid, ETSIM
Ciudad Universitaria S/N, 28040 – Madrid, Spain
{pabril,cvera,mta}@lst.tfo.upm.es,
hartaza@efronconsulting.com, jr.baragano@upm.es

Abstract. Nowadays, the new technological advances make possible to offer different services in a complete personalized way, covering the needs of heterogeneous user groups. In the case of elderly users, the acceptance of technology is a key aspect in their motivation to use certain services, and in consequence, the adaptability of the user interfaces is a critical requirement to achieve this goal. This paper presents the intelligent and adaptive user interfaces of a system devoted to offer AAL services, especially designed for increasing elderly users's acceptance, and developed as part of AmIVital project innovative technological platform.

Keywords: Ambient Intelligence (AmI), Ambient Assisted Living (AAL), Adaptative Interfaces, elderly users, multimodal interaction.

1 Introduction

Ambient Intelligence (AmI) paradigm aims to improve citizens' quality of life through an innovative utilization of ICT, where individuals are surrounded by networking and computing technology, unobtrusively embedded in their environment. User empowerment, efficient and distributed services, user-friendliness technology, and support for intelligent human interactions are key aspects of the AmI vision. The Ambient Assisted Living (AAL) concept applies this model to the provision of various services for elderly and people with special needs, helping them to better manage diverse situations in their daily lives. AAL services should include invisible, but always present technology; natural, simple and effortless interaction with users; device-independent and reactive user interfaces; and continuous, autonomous and ubiquitous service adaptation.

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 129–136, 2011.

This paper presents the intelligent and adaptive user interfaces developed as part of an innovative technological platform, which enables the creation of new AAL applications and services within a personal environment, for the care, monitoring and support of citizens with special needs in relation to their health and welfare status. The system has been developed within the framework of AmIVital research project [1], partially funded by the Spanish Ministry of Industry, Tourism and Trade.

AmIVital project introduces an adaptable User Interface Framework which is able to support the provision of a variety of complex AAL health and social services for elderly people (e.g. chronic obstructive pulmonary disease (COPD), heart failure (HF), and stroke assistance and management). The services have been built combining basic components: videoconferencing, vital signs and parameters monitoring (electrocardiogram, blood pressure, glucose, weight, and oxygen saturation), alerts, agenda, calendar and reminders. Each component, as well as the global services, includes a user interface that has been designed to be highly adaptive to the offered application, the specific user needs, and particular context situation.

2 Methodology

The methodology used for the design and development of AmIVital user interfaces, the "Interaction Framework" design method, comprises a combination of the principles stated in the User Centred Design [2] and Goal Oriented Design [3] methods. The whole process has followed an iterative approach that embraced three main phases: User Requirements Definition and Analysis, Framework Interaction Design and Development, and Design and Implementation of the User Interfaces.

In the first phase, User Requirements Definition and Analysis, all users' and functional requirements were elucidated and the concept of each service was defined. The work started with the examination of the available social and medical solutions to

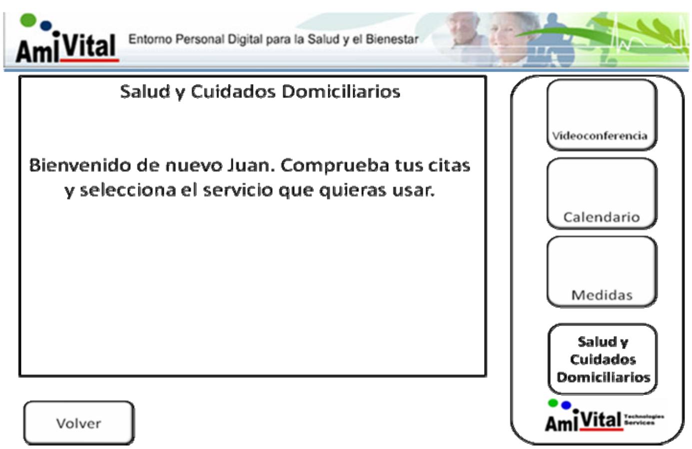


Fig. 1. This figure shows the functional blocks organization in the second version of AmIVital user interface prototype, taking into account that the user device is a tactile screen

assist the final target group, elderly chronic patients. Then, a selection of the different AAL services to be included into the system was done, following a Service-Oriented Modeling Analysis (SOMA) [4][6]. Finally, for each of the planned AAL services, particular user needs were identified, analyzed, and subsequently refined with the correspondent interaction requirements, such as simplicity and easiness to use, short curve of learnability, adaptation to personal routine, multimodality, guidance for use, and error prevention and recovery.

The core interaction framework was designed and developed in the second phase. This included a set of decision rules that, based on the previously identified user requirements and the environmental data acquired, define the best possible user interface at each time. At this stage, all functional elements of the interaction framework for each offered service were depicted (Fig. 1), as well as the key and secondary paths (Fig. 2).

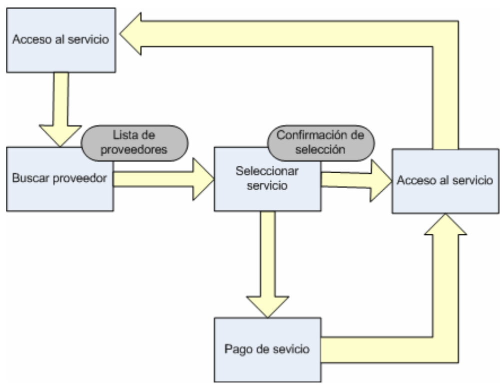


Fig. 2. This figure illustrates the key path of Videoconference service

Finally, the last phase involved the design and development of the actual user interfaces for each of the offered AAL services, including the number and position of the main functional elements of the graphical user interfaces (GUI), and multimodal functionalities such as the dialogs for voice interaction, and the vibration rules for haptic interaction. It also comprised the definition of the "look and feel", that is, the graphical aspect and the behavior of the data and functional elements in each application. The design took into account several accessibility guidelines especially focused on elderly user [5].

3 Results

Following the previously described approach, three different versions of the user interface were designed in an iterative way, including progressive and increasing complexity and functionality. Intermediate designs, obtained as outcome of each iteration step, have been tested and used to refine and improve the next one.

3.1 First Version Prototype

The first version of the user interface included the basic functionalities for the final services (agenda, reminders, videoconferencing and monitoring), and two basic environment adaptation rules, related with the user position and type of required monitoring. The main goal of this interface was to provide a rapid prototype that could be used for evaluating the acceptance of the basic set of the offered services.

The prototype interface was focused on implementing the basic functionalities of the final services, reducing the adaption possibilities to a few options (i.e. automatic selection of the type of monitoring for each user, and reminders displaying according to the user's behavior). This simple design facilitated the implementation of the interfaces, as the requirements were quite limited, and made possible the use of well known interaction devices, a television set and a remote control (Fig. 3).

Fig. 3. This figure illustrates two examples of the first prototype of AmIVital user interface, showing colors, visual elements and position of the interaction elements

Table 1. Validation of 1st prototype. Distribution of services per participants and age.

Participant	Service distribution	Participant's age
1	Videoconference with 2 vital signs monitoring, personal alarm and reminders	69
2	Videoconference with 4 vital signs monitoring, reminders, personal alarms	75
3	Videoconference with 3 vital signs monitoring, reminders, personal alarms	81
4	Videoconference with 3 vital signs monitoring, personal alarms, calendar	67
5	Videoconference with 3 vital signs monitoring, personal alarms, calendar	72

The resulting first prototype was tested with five users, in order to assess the usability and acceptability of the user interface design. Additionally, the ability of the design to introduce additional functionalities was also evaluated. Participants in the validations were senior citizens living independently in their own houses with different levels of chronic illness or disabilities (Table 1).

The obtained results were quite useful. Initially, four of the five elderly participants who had previously indicated their difficulties for controlling a computer based services were quite concerned and reluctant to use the system. However, their appreciation improved during the introduction of the system once they confirmed that the user interaction was based on a television set, a device which they were very familiar with.

All participants praised the simplicity of the design and the short learnability curve, and they also positively appreciated the introduction of visual elements or metaphors that facilitated the access to the different offered services. Nevertheless, in general, the system obtained low acceptance rates. This was mainly caused by the low level of adaptability and personalization of the application at that stage, as well as the reduced interaction possibilities provided by the remote control, especially critical when the number of offered services exceeded three and the time to access to a specific application suffered some seconds of delay.

3.2 Second Version Prototype

Based on the validation results of this first version, a second prototype of the interface was implemented, including extended functionalities and more user's adaptations than the previous one. Additionally, different options to be used as main interaction device were investigated. Aspects like easiness to use, naturalness and functionality were considered and, as a result, a touch screen was selected.

The new graphical interface design had the screen area divided in two main parts: the left side was the main interaction area showing the current service in use, while the right one (functions area) included all the buttons for accessing the different services and functionalities (Fig. 4). This design was decided in order to facilitate the interaction, based on the assumption that the majority of users are right-handed. In the function area the different applications were grouped in a personalized way, according to the importance for the user and the frequency of use (i.e. chronic patients would have on the top of the screen the services related to their illness, while other secondary applications, such as domotic control, would be placed at the bottom).

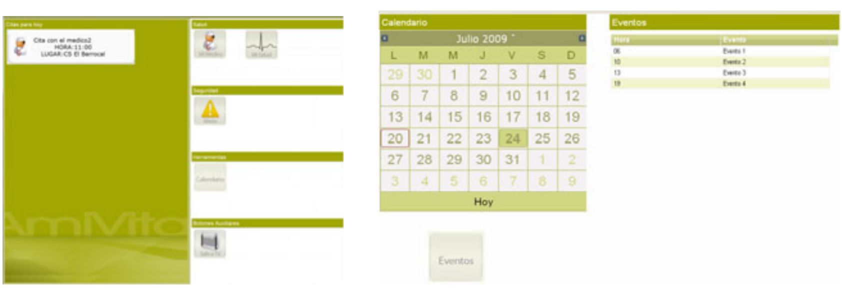


Fig. 4. This figure illustrates two examples of the second prototype of AmIVital user interface, showing colors, visual elements and position of the interaction elements

This second prototype of the interface was also validated with five end users, in order to assess the appropiate personalization of the services and applications according the user profile. Again, all participants were senior citizens living independently in their own houses with different levels of chronic illness or disabilities. The number of assessed services and their paticular combination were different for each participant, as it is shown in Table 2.

Table 2. Validation of 2nd prototype. Distribution of services per participants and age.

Participant	Service distribution	Participant's age
1	Videoconference with 2 vital signs monitoring, personal alarms, and reminders	65
2	Videoconference with 4 vital signs monitoring, domotic control, reminders, calendar, personal alarms	75
3	Videoconference with 3 vital signs monitoring, reminders, personal alarms	71
4	Videoconference 5 vital signs monitoring, domotic control, personal alarms, calendar	68
5	Videoconference with 5 vital signs monitoring, domotic control, personal alarms, calendar	76

The results showed that the acceptance of the system was very much affected by the number and distribution of elements on the screen. Users who had a reduced number of services in their interfaces gave very high rates to the system (participants 1 and 3), as well as to the selected device (the tactile screen), and the use of visual metaphors introduced to increase the usability and learnability of the interface. However, the acceptance decreased considerabily when the number of services was very high, forcing the interface to introduce scrolls or rotate menus (participants 2, 4 and 5). Addionally, results showed differences between young and old seniors. While old seniors considered the general design simple and easy to use, young seniors judged it as little attractive or low interactive. Finally, three participants appreciated the introduction of alternative interaction modalities, like sounds (participants 1, 3 and 5).

3.2 Third Version Prototype

The third and final version of the interface aimed at including multimodality access and creating a more attractive design, in order to increase the acceptance of the young elderly users and to provide alternative interaction methods. The design comprised the two well-accepted features of the previous prototypes: the TV-set navigation and the use of visual metaphors in the graphical user interface (GUI). Once again, a GUI was the basic interaction method, but in this case, the design also supported multiple interaction devices (i.e. PC, touch screen, TV set with remote control, Wii remote control, or web camera that detects the user's hand movements). Moreover, alternative interaction methods have been implemented into the final system, such as Text-to-Speech Technology (TTS), included as a support in some applications like the reminders, and Voice Recognition (ASR), as a secondary interaction method for the final users. These multimodality features were especially devoted to tackle the technophobia problems experienced by older seniors, by helping them to establish a natural dialog with the system and, therefore, increasing its acceptance (Fig. 5).

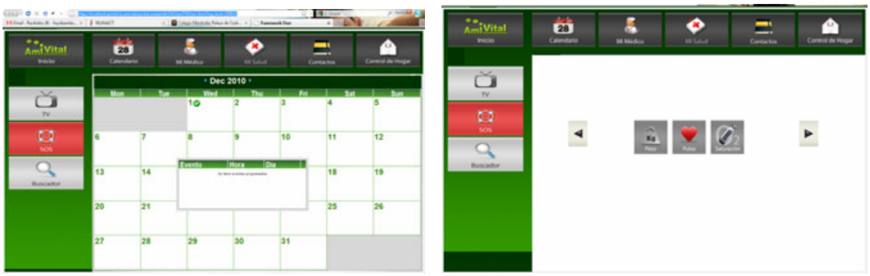


Fig. 5. This figure illustrates two examples of the third prototype of AmIVital user interface, showing colors, visual elements and position of the interaction elements

The validation tests of this final version have been especially focused on the multimodal and multi-device interaction aspects. The tests have taken place in the Ciami Living Lab [8] in UPM dependencies, involving six elderly users, living independently in their own houses (Table 3). Participants were free to use the graphical, voice recognition or text-to speech interactions, as well as different devices, in order to validate the diverse offered options.

Although the evaluation of the final interfaces has not finalized yet, preliminary results show high levels of acceptance of the multimodality functions among elderly, especially with the voice interaction. In the upcoming months, additional tests with end users have been planed, in order to assess all aspects of the developed user interfaces and AAL services.

Table 3. Validation of 3rd prototype. Distribution per participants and age.

Participant	Participant's age
1	65
2	78
3	69
4	72
5	71
6	73

4 Conclusions

AmIVital project final user interfaces have proved to be completely adaptable to the user needs and context, personalized in terms of desired services, and device independent. The design methodology utilized, the iterative development and intermediate validation procedure, and the involvement of end users in the different steps of the implementation, constitute an effective and efficient technique for developing AmI services' intelligent interaction. The validation of the resulting system with end users has shown high levels of acceptance, promising a quick and easy adoption the proposed AAL services once they are introduced into the market.

The final validation results are really promising. Participants who took place in the validation of the final prototype show high levels of acceptance and interest of the system. Their comments during the whole validation process were very useful to improve the interaction framework and a challenge for user interface designers.

As results of the process, the final interaction framework implementation has achieved high and accurate levels of adaptation and personalization that is expected to contribute to a better acceptance of the AAL services in the future and, therefore, to improve citizen's quality of life.

Acknowledgments. We would like to thank the whole AmIVital Project Consortium for their valuable contributions for the realization of this work. This work has been partially funded by the Spanish Ministry of Industry, Tourism and Trade, project AMIVITAL-CENIT. CENIT 2007-1010.

References

1. AmIVital project. Spanish Ministry of Industry, Tourism and Trade funded, Web page `http://amivital.ugr.es` (last access, February 2011)
2. ISO 13407. Human-Centred Design Processes for Interactive Systems, Geneva, Switzerland (1999)
3. ISO/IEC 9241. 200 series: Human system interaction processes (1998)
4. Bell, M.: Introduction to Service-Oriented Modeling. In: Service-Oriented Modeling: Service Analysis, Design, and Architecture, Wiley & Sons, Chichester (2008)
5. Selwyn, N.: The Information Aged: A Qualitative Study of Older Adults' Use of Information and Communications Technology. J. of Ageing Studies 18 (2004)
6. Newcomer, E., Lomow, G.: Understanding SOA with Web Services. Addison Wesley, Reading (2005)
7. Abril-Jiménez, P., Vera-Muñoz, C., Cabrera-Umpierrez, M.F., Naranjo, J.C.: Design Framework for Ambient Assisted Living Platforms. In: Stephanidis, C. (ed.) UAHCI 2009. LNCS, vol. 5615, pp. 139–142. Springer, Heidelberg (2009)
8. Montalvá, J.B., Fernández, M.M., León, A., Cabrera-Umpierrez, M.F., Arrendondo, M.T.: CIAmI: A Research center for AmI development and user experience evaluation. In: 1st ASK-IT International Conference, Nuremberg (2008)

Scratchable Devices: User-Friendly Programming for Household Appliances

Jordan Ash, Monica Babes, Gal Cohen, Sameen Jalal, Sam Lichtenberg, Michael Littman, Vukosi Marivate, Phillip Quiza, Blase Ur, and Emily Zhang

Rutgers University, Department of Computer Science, 110 Frelinghuysen Rd, New Brunswick, NJ – 08854, USA
{jordash,galcohen,samjalal,pquiza}@eden.rutgers.edu, {babes,mlittman,vukosi}@cs.rutgers.edu, blase@blaseur.com, {samlichten,1800.ehz.hang}@gmail.com

Abstract. Although household devices and home appliances function more and more as network-connected computers, they don't provide programming interfaces for the average user. We first identify the programming primitives and control structures necessary for the universal programming of devices. We then propose a mapping between the features necessary for the programming of devices and the existing functionality of Scratch, an educational programming language we use as a basic interface between the devices and the users. Using this modified version of the Scratch language, we demonstrate usage cases in which novice programmers can program appliances, increasing their functionality and ability to be customized. We also show how standardizing this programming paradigm can facilitate knowledge transfer to new devices. We conclude by discussing our experiences prototyping programmable appliances.

Keywords: educational programming, end-user programming, home automation, household devices, programming languages, scratch; ubiquitous computing, usability.

1 Introduction

From the Jetsons' futuristic home to the personalized temperature, music, and lighting systems in Bill Gates' estate [1], the concept of home automation intrigues and fascinates the average person. From a hardware perspective, home automation is already a reality as equipment for interconnecting devices is commercially available. With the arrival of an 'Internet of Things', a future proposed by major electronics manufacturers [2] in which every household device and appliance would have an IP address and be connected to the internet, capable hardware would become even more ubiquitous. However, the manner in which the average person might interact with and customize household devices has not been well studied.

One source of interest with home automation is that it is a mode of self-expression in addition to adding convenience and functionality to one's lifestyle. A home is important: it is, to many, a reflection of who they are and what they value. This human

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 137–146, 2011.

desire for customization propelled personalized ringtones into a multi-billion dollar industry [16]. It is a natural step to make a home's electronics and appliances into something personal, creative, and increasingly functional.

Existing home automation systems focus on the hardware for connecting devices, paying little attention to creating a usable and universal programming paradigm capable of controlling any type of appliance. The ability to fully customize a home requires the logical and control structures of an expressive programming language, whereas existing systems generally provide only a menu-based system for controlling appliances. Our goal is to enable users to write simple programs that add unique functionality, allow devices to communicate, and save time.

The overriding goal of our project is to enable the average person, somebody with no previous programming experience, to write programs for customizing their home appliances. To this end, we propose using a variant of the graphical programming language Scratch, designed for novice programmers [13], as the standard for programming home devices since Scratch is meant to be intuitive for first-time programmers.

In this paper, we lay out a comprehensive yet usable system for programming home devices. We outline the primitive operations necessary for such a programming system and show how the style and metaphors of the Scratch language directly support these operations. As initial support for our concept, we demonstrate usage cases building on these primitives. These compact example programs demonstrate novel, useful functionality for an automated home. We also discuss the construction of our prototype "Scratchable Devices."

2 Background and Related Work

We first explain the physical communication standards and programming methods of existing home automation systems and introduce the Scratch programming language.

2.1 Existing Home Automation Systems

Existing residential home automation systems primarily focus on communication protocols. X10 [17], one of the most widely used protocols, communicates over power lines. INSTEON, designed to provide added reliability over X10, enables communication over both power lines and wireless RF [14]. Protocols that are primarily wireless include Z-Wave [18]. Although these standards specify how home devices communicate, they do not prescribe usability standards or programming methodologies. Since our investigation focuses on the user interface, our proposal could be considered orthogonal to existing communication standards.

The control systems for existing home automation solutions range in complexity from touch screens to intricate (at times unintuitive) software packages that allow users to write programs in scripting languages. However, each of these approaches treats usability and functionality as mutually exclusive, in contrast to our proposed Scratchable Devices.

Hardware methods for controlling household devices include universal remote controls and wall-mounted touch screens [9]. Touch screen systems generally allow users to control devices, monitor their state, and set a device to change states at some later time. Software for smart phones and tablets, such as the iPhone and iPad, is also available from multiple manufacturers [4], [5].

Computer software, either local or web-based [8], is also commonly used to control home automation systems. Often, this software is controlled through a series of menus that closely mirror the functionality of touch screen interfaces. The user's ability to write their own programs is often limited to writing macros, as in the Active Home program [17]. These macros lack the control structures of full programming languages, such as iteration and conditional statements.

Some home automation systems do grant limited access to control structures. For instance, Zeus [15] allows users to define conditional statements, although this language lacks iteration and other hallmarks of fully featured languages. The PowerHome program provides customers with a full set of commands in the scripting language of their choice [11], but these languages are designed neither for novice users nor for the specifics of home-automation applications. None of these approaches successfully merges functionality with usability.

2.2 Scratch

Scratch, a graphical and interactive programming environment developed to foster children's excitement and interest in computer programming, has a number of features that make it suitable for first time programmers [13]. Scratch makes programming simple by representing statements as drag-and-drop blocks that snap together. These blocks include basic elements such as variables, conditionals, loops, inputs, and Booleans. Scratch is object-centered, where each object is called a sprite. Each sprite appears as an image whose graphical appearance, termed a costume, is under user control. All sprites appear on a stage, which is an area of the screen that displays each sprite in its current costume.

Users write event-driven scripts to control sprites. For instance, a user could create one sprite that looks like a cat and another sprite that looks like a dog. After choosing the cat sprite by clicking on it, the user could write a script that changes the cat's costume whenever the spacebar is pressed or make the cat meow if the dog is too close. Sprites can also communicate with each other through a broadcast mechanism.

To make household devices programmable to users without prior programming experience, the language they use must be simple to learn, intuitive, and present a uniform interface across different household devices. Because of Scratch's gentle learning curve and visually attractive, non-intimidating interface, it is used with great success in schools nationwide, from elementary through high school, and even in some introductory computer science courses at universities [6], [13].

3 Programming Primitives for Home Automation

To design a general-purpose programming language for household devices, we first propose a set of language-independent programming primitives that formalize users' interaction with their appliances. At a high level, we propose that these primitives be object-centered since users see each household device as a separate object. We also suggest that the language standardize the appearance and functionality of outputs and inputs. The output of each object will be state changes. Inputs to our language consist of sensors (on devices or free-standing), physical interaction with individual devices (such as button presses), and state queries. We additionally identify inter-device communication as an essential primitive. Our primitives are outlined in Fig. 1.

3.1 Outputs as State Changes

From a programming perspective, the most important property of a household device is its current state. The simplest devices have binary state; for example, a lamp can be either ‘on’ or ‘off’. More complicated devices may have many different states, but the user should still conceive of all manipulations to the device as a state change. For instance, a washing machine could have multiple modes of being ‘on’, such as being ‘on in *Permanent Press* mode’ or ‘on in *Bright Colors* mode’.

For both simple and complex devices, we propose that the primary output of a home automation programming language be changes of state on a per-device basis. Therefore, a simple device will have only a ‘Turn On’ block and a ‘Turn Off’ block. More complex devices would add a drop-down menu of possible modes to the ‘Turn On’ block. For instance, a washing machine's Turn On block could be edited via a drop-down menu to specify that the device must turn on in ‘*Permanent Press*’ mode.

3.2 Inputs

Many sensor readings will be available globally to all objects, including the time and date as well as environmental conditions. Sensor readings related to time include blocks such as ‘Time’ and ‘Is Workday’, where the latter block returns a Boolean value. A number of global sensors could measure the environment, such as the ‘Temperature’ or ‘Brightness’ in a particular room, allowing users to create a ‘Smart House’. These global sensors could be collocated with devices or placed independently throughout a home.

Novel Language Inputs	**Language**	**Novel Language Outputs**
Time [Time Is 12 : 11 PM ?]	Iteration	Basic Device State Changes [Turn On]
Environment Sensors [Temperature (F)]	Relational/ Logical Operators	Complex Device State Changes [Turn On In Permanent Press ... Mode]
Physical Interaction [HR Button Pressed?]	Conditionals	
	Delays/Timing	
State Queries [Device On?]	Randomness	
	Broadcast Communication	
	Multimedia - Audio	

Fig. 1. This table outlines the novel inputs and output primitives necessary for controlling household devices with a graphical language. Necessary control structures are also indicated.

Local inputs include both a user's physical interactions with a device and state queries. For instance, the action of a user pressing a button or turning a knob on a particular device would be represented, respectively, by a Boolean block and a block that returns a number. State queries, such as a 'Device Is On?' block and a washing machine's 'Device Is In [Cotton Cycle] State?' block, would return a Boolean value.

3.3 Communication between Devices

To permit rich interaction between household devices, a communication channel is necessary. Since the number of purpose-built communication channels between particular devices would increase exponentially with the number of devices, we propose that a user send messages as broadcasts into a cloud, with low-level communication transparent to the user. Each device would have an input block for receiving messages. Thus, a toaster could send the message 'Turn On Coffee Maker', and a coffee maker would have a corresponding script written by the user that listens for the broadcast 'Turn On Coffee Maker' and responds appropriately. Since messages are broadcasts, our proposed programming language enables one message to simultaneously communicate with many devices. Thus, the message 'Breakfast Time' could coordinate multiple devices working together every morning.

3.4 Control Structures

To present a user with the power to fully express his or her ideas, our language must contain the core control structures of modern programming languages. Iteration, conditionals, and both relational and logical operators are all essential. In addition, pseudo-randomness, multimedia functions, and the ability for objects to communicate all greatly increase the usefulness and expressiveness of this programming language.

3.5 A Device's Logic

A Scratchable Device should be sold with its logic and functionality programmed in our language and visible to the user. An interested user could view and modify the core operations of a device. For instance, a user could reprogram the buttons of a coffee maker to reflect the way he or she uses the device. Once the user writes a program, this program could reside on a centralized controller or on the device itself.

However, the logic that ensures the safety of a device should reside on the device itself and should operate independently of the programmable functions. For instance, a device with a heating element should be able to turn itself off based on time or temperature thresholds specified by the manufacturer in order to prevent a fire, regardless of whether a user's program is requesting the device to remain on.

4 Mapping Primitives to Scratch and Prototyping

The programming primitives we have outlined for Scratchable Devices can be mapped to the graphical programming language Scratch. Each device in a home automation system can be considered a sprite in Scratch, which functions as an object. The state of the device can be represented visually by a Scratch costume, and a user can view the current state of all the devices on the Scratch stage.

We have prototyped our programming paradigm using BYOB (Build Your Own Blocks) [10], a variant of Scratch in which we can create customized programming blocks. These blocks follow the style and structure of Scratch, with blocks such as 'Turn On' that appear as if they were native to Scratch.

5 Usage Cases

We believe that the ability to program home appliances will enable users to do *more*, extending the capabilities of these appliances beyond what is typical of devices today. In addition, users should be able to accomplish tasks *faster* by using programming rather than menus, and learn to use new devices *sooner* because of the unified interface across machines. To demonstrate the functionality of our proposed programming paradigm, we present the following usage examples.

5.1 'Clapper' for a Light

Our first usage case illustrates how novel yet useful features can be added to a device with short programs. In Fig. 2, we present a script that uses the computer's microphone and Scratch's 'loud?' Boolean, true when the microphone's input exceeds a threshold, to turn on or off a lamp. This code is placed as part of the lamp's sprite.

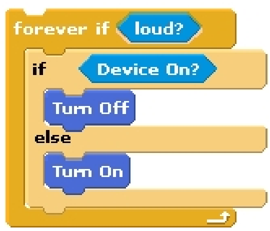


Fig. 2. This script allows any device to be turned on or off by clapping

Knowing how to program a 'Clapper' light makes it very easy to make a 'Clapper' fan or alarm. In fact, the Scratch code is identical, but would instead be placed within the sprite for the fan or alarm clock, respectively. Reusability of code makes introducing a new device to a household very simple. In contrast to current devices, the learning curve for new Scratchable Devices would be minimal.

5.2 Wake Up with Coffee

The Scratchable Devices paradigm makes inter-device communication seamless, whereas this sort of communication is complex or impossible in current systems. In Fig. 3(a), an alarm clock tells a coffee maker to turn on at 7:57 AM, waits three minutes, and sounds the alarm's buzzer. Thus, coffee will be ready when the user wakes up. The script in Fig. 3(a), written in the alarm clock's sprite in Scratch, broadcasts

(a) (b) (c)

Fig. 3.(a) An alarm clock requests coffee and sounds its buzzer afterwards, waking the user to a fresh-brewed pot. **(b)** The coffee maker responds to "make coffee" by switching on. **(c)** Each time the coffee maker goes on, it announces when the coffee is done.

the message 'make coffee'. The script in Fig. 3(b) must be included in the coffee maker's sprite in order to define what 'make coffee' means.

5.3 'Coffee Is Ready' with a Song

Scratchable Devices can also be customized with rich multimedia, similar to cell phone ringtones. For instance, users can program their coffee maker to play their favorite song when their coffee is ready, as shown in Fig. 3(c).

5.4 Activity Simulator for Vacations

Elements common to most programming languages can also add functionality to homes. Fig. 4 depicts a program that uses a pseudorandom generator to turn a house light on and off randomly after 6 PM, deterring potential thieves while a user is away on vacation. Pseudorandomness, while simple to employ in a language like Scratch, would be difficult to implement without a fully-featured programming language.

Fig. 4. A house light is turned on and off during evening hours to deter would-be robbers

5.5 Setting Alarm Time

In existing systems, changing a clock's alarm time from 9:00am to 8:45am would generally require around 68 button presses to advance both the minute and hour. However, small changes in Scratchable Devices can be effected by editing a text box, or the user can reprogram existing buttons for common operations. This same mechanism works across devices, resulting in knowledge transfer.

More usage examples can be found on our scratchabledevices.com website.

6 Prototype Construction

We have constructed prototype Scratchable Devices, including lights, fans, and alarm clocks. While our prototypes demonstrate the feasibility of this programming paradigm and provide an upper bound on the hardware costs, existing or future hardware systems for home automation could just as well form the hardware layer of Scratchable Devices. Homemade Scratchable Devices could similarly exist. The primary contribution of this paper is the *user experience* in the household devices domain rather than a specification for hardware-level control.

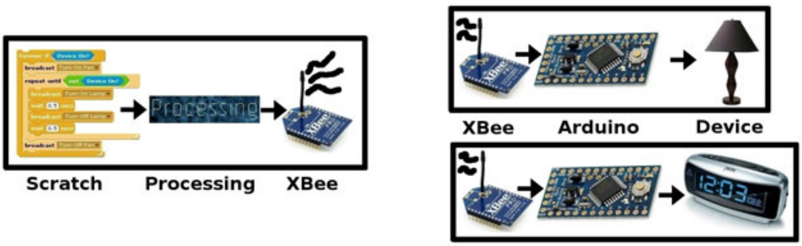


Fig. 5. Diagram of our prototype system, which uses the Processing language to pass commands from Scratch wirelessly (using XBee) to devices controlled by microcontrollers

On the user experience level, our Scratchable Devices prototypes have been implemented using the BYOB (Build Your Own Blocks) variant of Scratch. We chose BYOB because a number of our proposed programming primitives, including the `Turn On' block for each device and our use of time as a global sensor, required new types of Scratch blocks.

To allow BYOB to communicate with physical devices, we have used the Processing language as a translator [12]. A program in the Processing language runs in the background at the same time as Scratch, detecting when a variable in Scratch has been changed. Each Scratch sprite (household device) has one or more variables indicating its state. Each physical device has a globally unique 3-byte ID, and each corresponding sprite's variables begin with that ID. Our program in Processing takes messages received from Scratch and broadcasts them without modification over an XBee 802.15.4 wireless module.

Every Scratchable Device, such as a lamp or fan, also has an XBee 802.15.4 wireless module connected to an Arduino hobbyist microcontroller [3]. Since communication takes place on a broadcast medium, all devices receive all messages. In a commercial implementation, it would be wise to define a bootstrapping procedure in which newly purchased household devices inform the local Scratch instance of their device ID and feature set. That device would automatically appear as a sprite (object) in the Scratch instance.

When a Scratchable Device receives a message addressed to its ID or a broadcast message, such as 'Breakfast Time', it is programmed to read the message payload and change the state of the device accordingly. We have reverse-engineered a number of household appliances and inserted relays, LCD screens, and additional mechanisms that are controlled by Arduino microcontrollers. We have also attached small circuits to the buttons of the devices so that the microcontrollers can detect physical button presses and pass this information to Scratch, via Processing.

7 Conclusion and Future Work

We presented Scratchable Devices, a programming paradigm for household devices that allows the average person to program his or her appliances using a graphical programming language. The logical and control structures of a programming language provide additional functionality to household devices and potentially simplify existing tasks. We presented the programming primitives necessary for interaction with household devices and showed how the graphical, educational programming language Scratch maps directly to these primitives. We provided a short list of simple usage cases in our programming paradigm to support our claims.

In the near future, we will conduct user studies to verify the usability of the system compared to standard interfaces. Using this kind of formative design-evaluation [7], our interface will be made entirely with the average user in mind. We are particularly interested in verifying that users with no programming experience can, with minimum time and effort, learn to program devices in ways that add novel operations. We are also interested in any cognitive dissonance between our paradigm and the conception the users have of their household devices.

In the long run, we want to encourage electronics manufacturers to employ a version of our programming paradigm as an industry standard. Both similar and dissimilar devices should all use the same programming protocol. This consistency throughout the home would make for a simplified, richer user experience.

Acknowledgments. The authors thank Roy Cohen, Matt Continisio, Steven Fisher, Rich Katz, Molly Littman, Lisa Littman, Luis Piloto, Amanda Rumsey, and Raheem Scott-Griffith for their contributing ideas to the design and construction of prototype Scratchable Devices. This material is based upon work supported by the National Science Foundation under Grant No. NSF IIS-0713148 (REU supplement).

References

1. Allon, F.: Space, Media Flows and 'smart'living in the Absolute Present. MediaSpace: Place, Scale, and Culture in a Media Age, 253 (2004)
2. Vance, A.: You Too can Join the Internet of Things, `http://bits.Blogs.Nytimes.com/2010/09/20/you-Too-can-Join-the-Internet-of-Things/`
3. Banzi, M.: Getting started with arduino. Make Books (2008)
4. Control 4: My Home - iPad, `http://www.control4.com/residential/products/mobile/myhome-Ipad/`
5. Creston Electronics Inc.: Apple & Creston. iPhone, iPod and iPad Mobile Control, `http://www.Crestron.com/solutions/apple_mac_iphone_ipod_ipad_control/`

6. DesJardins, M., Littman, M.: Broadening Student Enthusiasm for Computer Science with a Great Insights Course, pp. 157–161 (2010)
7. Egan, D.E., Remde, J.R., Gomez, L.M., et al.: Formative Design Evaluation of Superbook. ACM Transactions on Information Systems (TOIS) 7, 30–57 (1989)
8. Hawking Technologies Inc: HomeRemote Pro Home Automation Starter Kit, http://www.Hawkingtech.Com/
9. HomeSeer Technologies LLC: Home Automation Systems, http://store.Homeseer.com/store/Touchscreen-User-Interface-Plug-Ins-C98.Aspx
10. Monig, J., Harvey, B.: Build Your Own Blocks. University of California, Berkeley (2010)
11. PowerHome Automation LLC: PowerHome Automation, http://www.Hawkingtech.Com/
12. Reas, C., Fry, B.: Processing: A programming handbook for visual designers and artists. The MIT Press, Cambridge (2007)
13. Resnick, M., Maloney, J., Monroy-Hernández, A., et al.: Scratch: Programming for all. Commun. ACM 52, 60–67 (2009)
14. SmartLabs® Inc: How INSTEON Works, http://www.Insteon.net/about-Howitworks.Html
15. Trax Softworks Inc.: Zeus Description and Features, http://www.Hawkingtech.Com/
16. Wagner, A.: The Mobile Storefront. Let Your Fingers do the Shopping, 45 (2005)
17. X10.com: Standard and Extended X10 Code Protocol, http://software.x10.com/pub/manuals/xtdcode.Pdf
18. Z-Wave Alliance: About Z-Wave, http://www.z-Wave.com/modules/AboutZ-Wave/

Passive Identification and Control of Arbitrary Devices in Smart Environments

Andreas Braun and Felix Kamieth

Fraunhofer Institute for Computer Graphics Research - IGD, Darmstadt, Germany
{andreas.braun,felix.kamieth}@igd.fraunhofer.de

Abstract. Modern smart environments are comprised of multiple interconnected appliances controlled by a central system. Pointing at devices in order to control them is an intuitive way of interaction, often unconsciously performed when switching TV stations with an infrared remote, even though it is usually not required. However, only a limited number of devices have the required facilities for this kind of interaction since it does require attaching transceivers and often results in the necessity to use multiple remote controls. We propose a system giving a user the ability to intuitively control arbitrary devices in smart environments by identifying the appliance an interaction device is pointed at and providing means to manipulate these. The system is based on identifying the position and orientation of said interaction device, registering these values to a virtual representation of the physical environment, which is used to identify the selected appliance. We have created a prototype interaction device that manipulates the environment using gesture-based interaction.

Keywords: Pointing device, gesture-based interaction, smart environments.

1 Introduction

Smart environments are comprised of various interconnected devices, e.g. lighting and heating. They can be roughly grouped into sensors, actuators, output devices and computational elements. Typically those devices are heterogeneous concerning their purpose, interaction metaphors and technological basis. Modern home automation and middleware systems are able to abstract the technological basis to a certain extent but heterogeneous input modalities remain an issue. Those systems are usually controlled by a central server and the user can interact with it using remote controls or network-enabled systems like smartphones or tablets. Gestural interaction with a system is a natural, intuitive way for controlling programs that enjoyed considerable market success in recent years with devices like the Nintendo Wii. We have developed a methodology that will allow appliances in smart environments to be controlled using sensor-equipped interaction devices. Based on determining the position and orientation of an input device it is possible to register those values into a virtual representation of the smart environment and identify an appliance the input device is currently pointing at. Based on this methodology we have created a first prototype acting as proof-of-concept device for this kind of user-system interaction. This prototype is

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 147–154, 2011.

focusing on the sensor-equipped interaction device and is integrated to a simulated smart environment instead of a real one. It allows the user to modify the virtual appliances performing gestures with the actual prototype.

2 Related Work

Virtual representations of smart environments have been commonly used in research, e.g. as user interface metaphor. A three-dimensional model of the smart environment is often directly used as graphical user interface that either allow direct manipulation of appliances [1] or is augmented with classical user interface items [2], distinguishing between selection and manipulation. Another research topic has been investigating virtual representations of smart environments to allow rapid prototyping and user-centered testing of apartments and buildings in the virtual realm [3], e.g. in the context of ambient assisted living, creating apartments that are suited for persons with physical impairments.

There are various different principles of indoor localization available. A common method is using triangulation or multilateration of radio-frequency signals, whereas a device is tracked by using geometrical information of several senders/receivers and time-of-flight data [4]. However precision is usually limited and therefor this method is mostly suited for detecting the occupation of a room, not the exact location within it. Optical tracking is another method for indoor localization that is using feature extraction in combination with camera parameters and known positions of several cameras to calculate the position of an object in 3D coordinates, e.g. using a stereo camera system and blob detection to track multiple persons [5]. An unobtrusive but technically complex solution is using active floor systems. These are based on pressure or presence sensing equipment in a floor that is able to determine the position of one or more persons. SensFloor [6] is such a system, relying on capacitive sensors that are embedded into the floor covering and communicate wirelessly with a central system.

Research regarding pointing at devices in smart environments has resulted in the development of the gesturePen [7], a line-of-sight based system that is able to identify appliances. It allows interaction with devices that are equipped with wireless RF-tags. Therefore it allows manipulation based on pointing without the usage of a virtual environment but requires every controllable appliance to be equipped with additional hardware.

3 Appliance Identification

As previously mentioned the appliance identification method is based on two premises. The smart environment needs the ability to reliably track position and orientation of the interaction device and the system needs a virtual representation of the smart environment, where all appliances and their respective positions are modeled.

3.1 Methodology

The basic idea of this method is to first determine position and orientation of an interaction device in the smart environment. These values are registered into the virtual representation of the latter, making it possible to apply discovery methods to find the appliance the device is pointed at in the real world. The system then is able to give commands to the connected appliance.

The general method can be grouped into four main steps leading from detecting the interaction device to sending commands to the devices placed in the smart environment. The first step is, as noted in the prerequisites, to provide means for the smart environment to register the position and orientation of the interaction device. The second step is to register orientation and position into the virtual representation of the environment. In the third phase we use this data to identify the appliance via discovery methods in the virtual environment. Finally the performed user interaction is registered and the selected device is controlled accordingly. The whole process is shown in Fig. 1. A more detailed description of the different steps and how they have been realized in the created prototype is shown in the following subchapters.

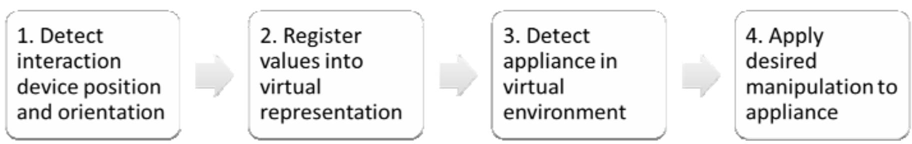


Fig. 1. Process to manipulate appliances via interaction device

3.2 Interaction Device Position and Orientation Detection

The position and orientation of any object can be described using six degrees-of-freedom. Most commonly Cartesian coordinates are used to define the position and Tait-Bryan angles (roll, pitch and yaw) to describe orientation. These values can be detected combined or independently of each other. Our proposed system uses an independent approach, whereas the position is determined using an indoor localization system based on an active floor, having ability to sense presence and position of a human being and an interaction device equipped with sensors to determine orientation of the device. The position of the device has to be tracked with a certain precision in order to achieve sufficient accuracy of the appliance selection algorithm. RF-based tracking of devices using time-of-flight data to track the position of a sender is usually suffering from an error in the range of several meters [9].

There are various options to determine the device position within the smart environment. It is possible to use optical markers on the interaction device a camera can detect and calculate position from [8]. Another option is to use a generic indoor localization method and estimate the device position using average body parameters, e.g. shoulder height and arm length. This will require factoring in an amount of uncertainty that will have to be considered when designing the identification of the virtual appliances and the feedback output devices give to the user. A common method to determine the orientation of a device is using an IMU (intertia-measurement unit) that is comprised of accelerometers, gyroscopes and magnetometers, having the ability to determine orientation and velocity of an object. These sensor units are nowadays installed in many multimedia devices, ranging from game console controllers to smartphones. The velocity data can additionally be used to improve the position detection of the device, using dead-reckoning[1] to estimate device location changes, thus creating more samples for the localization.

3.3 Registration in Virtual Environment Representation

Transforming position and orientation to the virtual representation is trivial using simple geometric transformations, based on the system knowledge about the real environment. We use the virtual position to project a ray along the virtual orientation and use ray intersection methods with the bounding volumes (BV) of the modeled appliances. There are various strategies to modify this ray intersection. The obvious solution is to select the first bounding volume that is hit. Other options include using metadata to prioritize important appliances or give a user the option to select between all detected objects on the path of the ray via the user interface.

Another option to modify this process is to use virtual appliances. These are regions within the virtual representation that have been assigned to a certain appliance without physical reference. This is useful to model very small appliances like lighting switches or hidden appliances like heating. For example now in the virtual representation a region at the ceiling can be assigned to manipulate the lighting, or a region on the floor can be assigned to heating control. In either the case the user needs feedback from the user interface to successfully select these strictly virtual regions.

Coping with very small or distant appliances resulting in small BVs, as well as occluded appliances, causing ambiguous intersections have to be taken into account using various strategies of modifying BVs in the virtual realm. Examples are distance-based scaling or morphing the shapes of occluded BVs. Similar strategies can be applied to increase the amount of used space if the appliances are sparsely distributed in the smart environment.

4 Interaction Device Prototype

We have created a first proof of concept interaction device (Fig. 2), equipped with a minimal IMU system and gestural interaction modality. The system is based on the Arduino[2] platform and uses a three-axis accelerometer and a single-axis magnetometer to determine orientation. The accelerometers are used to determine roll and pitch of the device. The magnetometer detects yaw relative to the earth magnetic field. The Arduino board is additionally equipped with a wireless communication chip based on the ZigBee protocol that is used to transfer generated data to the system. Signal processing is implemented on the Arduino microcontroller. This includes calibration of the sensors and scaling of acquired data to generate radian values that can be used in the subsequent processing steps.

This prototype possesses no facilities that aid the localization (with the exception of supporting dead reckoning) and therefore has to rely on the smart environment to track the position of the user and extrapolate device position. User-interaction is

---

[1] Method to estimate the location relative to a known position, e.g. by integrating over velocity values generated by an IMU.

[2] `www.arduino.cc` - open-source electronics prototyping platform.

realized using a single button and interpretation of the accelerometer data using the wiigee[3] library for Java that supports learning gestures based on collected training data using machine learning methods. Concerning the prototype the appliance selection and appliance manipulation through gestures is separated. If the attached button is pushed the device will stop appliance identification, assuming the most recently selected as desired and subsequently use accelerometer data to detect performed gestures that activate certain events. A small GUI-based program is supporting the learning process and stores gestures into readable files. These can be freely associated to different manipulations of the various appliances. For example a generic left-to-right gesture could be used to switch channels if television is selected or to increase brightness if the interaction device is pointing at the room lighting.

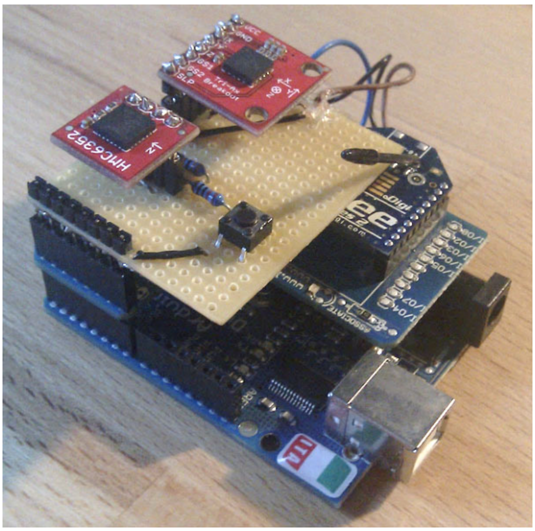


Fig. 2. Image of the Arduino-based prototype interaction device with accelerometers, magnetometer and button visible on the top board

5 Simulated Environment

In order to realize this proof-of-concept we have established a minimum simulated environment equipped with various virtual appliances that change their state based on performed gestures. At this point this virtual environment is not based on a real room and not connected to any indoor localization methods. However, it is possible to manually control the position of the person to test the device identification from various points within the simulated environment. An example scene of this virtual environment is shown in fig. 3. The user is in a virtual room and can control four appliances, two different windows, a TV set and a radio. In this scene the virtual person points at the TV, causing the associated BV to be highlighted. This way the user

[3] `www.wigee.org` - created for interfacing with Nintendo Wii Remote.

is getting a concise visual feedback, allowing him to easily determine that he has successfully selected the desired device. Now various gestures can be performed, which in this prototype state are limited to changing textures of the appliances. The system can be connected to a home automation system or middleware at a later point in time, to allow manipulation of actual appliances in a real smart environment, where instead of mapping detected gesture events to the virtual devices they can be sent to the home automation system controlling the actual device.

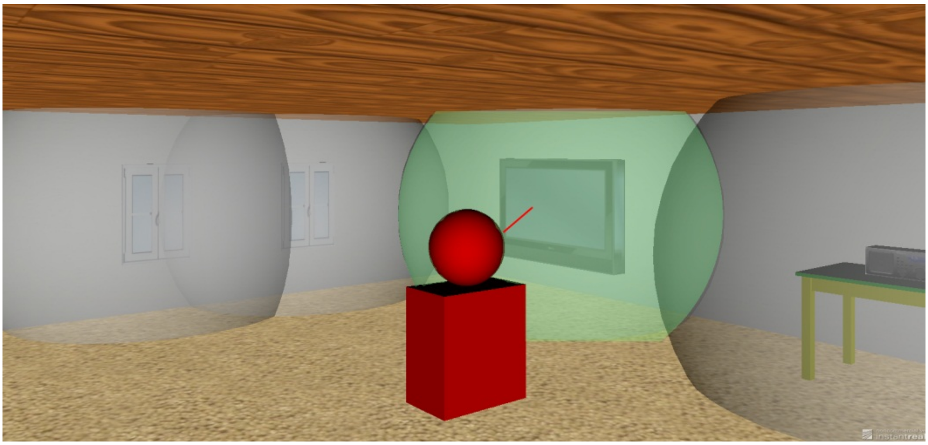


Fig. 3. Simulated environment with ray projected from the manually controlled person, bounding volumes of all integrated appliances and the green-tinted selected bounding volume

The acquired orientation data is registered to the simulated environment and used to cast a ray in the desired direction. We use the trivial method of selecting the first device, whose bounding volume is intersected and use gestures to manipulate the appliances in the simulated environment. Coming back to the process initially modeled the realization of our proof-of-concept system can be combined as shown in fig. 4.

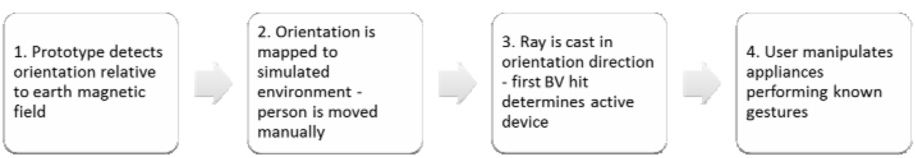


Fig. 4. Passive identification and control process as realized in the current prototype

6 Limitations

The whole system is based on the prerequisites that the smart environment needs to reliably track position and orientation of the desired interaction device and the system needs a virtual representation of the environment with all appliances modeled. This limits the system to static environments. If appliances are in the environment that are moved regularly, strategies have to be involved that compensate this limitation, e.g. providing simple interfaces that allow the user to reconfigure the environment.

The IMU is using magnetometers relying on the earth magnetic field to detect yaw. In home environments there may be various other sources of electromagnetic or static magnetic fields that will disturb these measurements and in worst case prevent the method from working altogether. It may be viable to include a Hall-sensor that can detect external magnetic field to let the user know why the interaction is interrupted, for example through GUI notifications or audio signals, alternatively it is possible to use on-the-fly calibration methods to allow for improved results in changed environments.

Currently differences in body posture and height are not explicitly compensated in the calculation process. It is assumed that appropriately sized bounding volumes and clear visual feedback like highlighting the selected appliance will allow the user to perform the interaction with sufficient precision. However this will have to be confirmed in a planned future evaluation of a completely implemented system in a real smart environment.

7 Conclusion and Future Work

By relying on a capable smart environment control unit and sensor-equipped interaction devices the complex task of identifying appliances the user points at is reduced to two components, allowing for a flexible solution that does not require the controlled appliances to feature active receivers or transmitters communicating with the interaction device. We have created a first realization of the sensor-equipped interaction as proof-of-concept relying on a minimal IMU system and a demonstration virtual representation that is not connected to a home-automation system.

As future work we intend to connect the system to our living lab that features various appliances connected to a home automation system. In combination with active floor indoor localization technologies the whole process of passively identifying arbitrary device can be realized in an actual smart environment. We will perform an evaluation of the complete system, focusing on the effects of different body parameters and pointing habits, in order to further improve the methodology.

Other points of interest include testing commonly available devices including IMU-like systems, e.g. smartphones and tablets. The main interest is to evaluate whether their integrated sensor units are sufficiently precise to perform the device identification with otherwise unchanged parameters. We plan to test other methods of indoor localization, e.g. optical tracking of a marker attached to the interaction device or depth imaging object recognition that should allow tracking of both position and orientation of an interaction device, eventually allowing the user himself to act as input device. These methods should be compared with regard to technical viability, performance and user acceptance. A final point of interest is the optimization of the appliance identification in the virtual representation using the BV optimization, hinted at in section 3.3 that should improve the method regarding the identification of small or distant appliances, view-point occlusion and sparsely distributed devices in the smart environment.

References

1. Borodulkin, L., Ruser, H., Tränkler, H.-R.: 3D virtual smart home user interface. In: Proceedings of the IEEE International Symposium on Virtual and Intelligent Measurement Systems, Girdwood, AK (2002)
2. Shirehjini, A.A.N.: A novel interaction metaphor for personal environment control: Direct manipulation of physical environment based on 3d visualization. In: Computers & Graphics, Special Issue on Pervasive Computing and Ambient Intelligence, vol. 28, pp. 667–675. Elsevier Science, Amsterdam (2004)
3. Kamieth, F., Roeder, P., Hellenschmidt, M.: Accelerated User-Centred Design of AAL solutions using simulated environments. In: Proceedings Ambient Assisted Living, January 27-28. VDE Verlag GmbH, Berlin (2009)
4. Lim, H., Kung, L., Hou, J.C., Luo, H.: Zero-Configuration, Robust Indoor Localization: Theory and Experimentation. In: Proceedings of IEEE INFOCOM, pp. 123–125 (2006)
5. Krumm, J., Harris, S., Meyers, B., Brumitt, B., Hale, M., Shafer, S.: Multi-camera multi-person tracking for easyliving. In: Proceedings of the Third IEEE International Work-shop on Visual Surveillance (2000)
6. Joint Project SensFloor, http://www.sensfloor.de/
7. Swindells, C., Inkpen, K., Dill, J., Tory, M.: That one there! Gesture for on-demand device identity. In: Proceedings UIST, Paris (2002)
8. Welch, G., Bishop, G., Vicci, L., Brumback, S., Keller, K., Colucci, D.: The HiBall tracker: High-performance wide-area tracking for virtual and augmented environments. In: Proceedings ACM Symposium on Virtual Reality Software and Technology (December 1999)
9. Taneja, S., Akcamete, A., Akinci, B., Garrett, J.H., East, E.W., Soibelman, L.: Analysis of three indoor localization technologies to support facility management field activities. In: Proceedings of the International Conference on Computing in Civil and Building Engineering, Nottingham, UK (2010)

Studying the Role of Interactivity in Museums: Designing and Comparing Multimedia Installations

Pedro Campos[1,3], Miguel Campos[2], João Pestana[2], and Joaquim Jorge[3]

[1] University of Madeira and Madeira Interactive Technologies Institute
9000-390 Funchal, Portugal
[2] WowSystems Inc., R. Mary Jane Wilson 23 G, 9050-446 Funchal, Portugal
[3] VIMMI Group, Visualization and Intelligent Multimodal Interfaces, INESC-ID
R. Alves Redol 9, 1000-029 Lisboa, Portugal
pcampos@uma.pt, miguel.campos@wowsystems.pt,
joao.pestana@wowsystems.pt, jaj@inesc-id.pt

Abstract. Interactive installations for museums are a particular kind of interactive systems, the design of which has been the subject of several research studies. However, the aspects of a rich, cultural experience are easily overlooked in a technologically driven system design and there are few studies that actually compare the role of different interaction styles (such as *touching* versus *walking*) on the museums visitor's experience. We present our experience of designing a cultural interactive multimedia exhibition, comprised of four sensor-based interactive installations, and two non-interactive installations. Our results were organized around usability problems detected, social interaction issues and differences between interaction styles, and suggest that the most enjoyable installations are those which facilitate collaborative activities as well as those making a creative use of sensor-based technology.

Keywords: Interactive installations, sensor-based interfaces, public settings, cultural exhibitions.

1 Introduction

The use of sensor-based interactive installations, in particular installations involving infrared motion sensors as well as cameras coupled with real time video processing algorithms, have been receiving considerable interest both from industry and academia [1, 2, 3, 5]. During the design and evaluation of interactive exhibitions, much can be learned about interaction design for public settings like these.

We describe our own experience designing and evaluating an interactive exhibition, which featured four different interaction styles to control digital contents: touching, walking over, waving and page-flipping. Our design approach was tailored to the exhibition's contents and makes a creative use of sensor-based technology, with the explicit goal of reducing the distance between visitors and cultural heritage [7].

The traditional view of culture-related exhibitions is that the visiting space should be well conceived in terms of volumetric layout so that it can essentially serve as a

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 155–164, 2011.

quiet contemplation space. Current approaches are still based essentially in static paper-based displays that are manipulated by artists who try to make them more appealing. However, with the advent of novel technologies, particularly multimedia projections and sensor-based installations, the museum directions and the exhibition's cultural and artistic directions are starting to embrace new digital media as effective ways of approaching people to cultural heritage, as opposed to considering those media as a menace to traditional means of cultural dissemination. Our experience suggests that artists have the potential to provide novel, creative uses to technology, and the opposite as well: interaction designers and technologists have the potential to provide artists with techniques which effectively enhance their portfolio.

The remainder of this paper is organized as follows: the next section describes related work with a particular emphasis on research approaches to add interactivity to cultural heritage exhibitions and museums. Next, we describe the interactive installations designed, as well as the interaction styles employed in our exhibition, title "Cultural Tourism". We move on to describe the evaluation approach and results, drawing some conclusions, which were organized around usability problems detected, social interaction among visitors and differences between interaction styles. Section "Conclusions and Future Work" outlines new avenues of research for this field.

2 Related Work

Technology today provides exciting new possibilities to approach museum visitors to culture and heritage. And many art museums are struggling to identify innovative approaches to engage new audiences in their exhibitions.

Danks and colleagues [1] refer the focus shift towards using interactive artifacts to enhance the visiting experience, which contrasts with the traditional approach, centered on the museum's collection, display and storage of objects. They argue that today's museum visitor expects to learn while also having fun at the same time, therefore interactive storytelling and gaming have a great potential to improve modern museum's experiences [1].

Several experiences have been conducted to study how visitors experience novel interaction styles within museums and science centers. Explore@Bristol, for instance, was an interactive science museum, which was studied to analyze six of its exhibits according to three dimensions: *Drama-Sensation*, *Challenge-Self Expression* and *Social* [5]. The exhibition titled a "Walk in the Wired Woods" illustrates how to design an engaging experience through context-sensitive media and interaction. The visitors were invited to take a walk in which they were automatically presented with audio content appropriate to their physical location [6]. Other interesting studies have been performed, regarding novel interaction styles and schemes, applied both to leisure and educational activities. Such examples include "The Fire and the Mountain" exhibition, held in 2006 at the Civic Museum of Como, Italy [4] and the "Listen Reader" from Xerox PARC, an innovative and engaging reading experience installed in three different museums over a six-month exhibition period [3].

Some researchers have devoted effort into studying interactive installations using mixed-reality [8], in the context of art museums. Expressing the formal aspects of the original artworks, the interactive installations allowed visitors to explore specific conceptual themes through their interactions. Sometimes, researchers also exploit an augmentative approach, adding interactive elements to the displays and artworks of the exhibition [9]. Taking great care to ensure that the installations meld seamlessly into the setting is considered very important, so that visitors don't face the interactive installations as a kind of "computer section" of the museum or exhibition.

Experiments on augmenting art museums with interactive technology have also been documented. For instance, Terrenghi and Zimmermann [11] introduce the notion of 3D sound in headphones for an art museum, providing the user with a contextual and spatial audio guide. This technology is an advanced version of more traditional audio guides. However, the approach still provides only an individual and detached experience, since no conversation is possible while listening to the audio. This communication approach does not disturb the experience of purely visual artworks, however it would be difficult to combine with artworks that contain sound by itself.

The use of abstraction and motion in the design of social interfaces – for which the interactive cultural exhibitions are a special case – has also been explored [10]. Particularly useful for our research was the concept of *perceptual causality*, which suggests that simple displays in motion can evoke high-level social and emotional content.

A particular requirement for designing our interactive installations was the goal of captivating children. Mark Prensky stated: "Our students have changed radically. Today's students are no longer the people our educational system was designed to teach." [12]. The same philosophy applies to the younger visitors of a museum. It is not enough to simply add technology to some of the exhibition's modules, designers need novel approaches to reduce the distance between children and cultural heritage and new interactive techniques to stimulate active participation, involvement and learning by children visiting a museum, through ubiquitous computer technology. A good example is provided by [13], who undertook a systematic design process, which involved exploring Scenario-Based Design, Design-Based Research and a number of technology probes. The evaluation took a case-based approach using video recording and post hoc analysis of the activities, discussion, reaction, and questioning by the children who visited their exhibition, both as individual participants and in interactive groups. The data derived from these video recordings was analyzed in the context of eight design themes, which informed the development of the novel, computer-augmented museum exhibition.

When studying visitors' interaction with museum digital installations, the problem of evaluation arises inevitably, since it's neither easy nor clear to establish the right set of methods in order to draw credible conclusions. Hornecker and Stifter [2], reporting on the evaluation of a digitally augmented exhibition on the history of modern media, based their conclusions on (i) logfiles' analysis, (ii) interviews and (iii) observation in the museum. We also followed this approach as a means to obtain a better understanding of the interactive installations role in the exhibition, described in the next section.

3 The "Cultural Tourism" Exhibition

Museums, as well as the cultural exhibitions they organize and create, are a heterogeneous set of institutions whose original twin functions of scholarship and education, once inseparable, but subsequently divorced, are being reunited by digital technologies. These technologies not only facilitate and/or accelerate long-established learning tasks, but, critically, they allow activities that would otherwise be impossible. This includes new approaches to learning by different audiences and for different purposes.

We designed a set of sensor-based installations in a cultural exhibition organized by the Direction of Cultural Affairs, which aimed at showing the visitor the cultural richness that formed the streets of Funchal (Portugal). The concepts of the exhibition revolved around promoting awareness about, and fostering a better understanding of, the cultural tourism that can be performed by simply walking through strategic streets and watching certain buildings, sites, and heritage. To better complement the exhibition's traditional large-format printed panels, the organizers wanted to have the interactivity factor as a means to add value to the visitor's experience.

The final set of interactive installations was comprised of: (i) a virtual book that could be browsed by simple page-flipping gestures performed in mid-air; (ii) an interactive floor that illustrated the evolution of the transportation means along the years; (iii) an interactive timeline using a touch-screen and (iv) a panel with projected images that would change through waving. The final set of dynamic but not interactive installations was comprised of (i) an immersive room featuring a 360-degree projection of traditional fireworks and (ii) a ceiling projection.

Fig. 1. The first interactive projection, which is "activated" through waving.

Figure 1 illustrates the beginning of the visit. A panel suspended from the ceiling is the projection surface for an interactive installation aimed at displaying ancient photos, which depict the evolution of the transportation means, and the cultural habits associated to it. Pictures were activated by simple waving gestures in front of the panel, as shown in Figure 1. The artists wanted to align the static panels, which they had in mind for displaying other contents, with the multimedia installations, so this installation's design was constrained by that goal. It is interesting to note that this is the sort of aspects that HCI researchers and interaction designers wouldn't think about during the design of such an interactive installation: the material of the medium (the projection surface), as well as its placement. We found this to be an important success factor in the exhibition.

Moving on, visitors could find an interactive historical timeline, which consisted of a conventional 42-inch touch screen. Cultural directors usually employ timelines in books, panels and lectures. It's the most common format for displaying historical events and dates. The choice of the interaction style and the whole design of this installation were, therefore, immediate. From a research point of view, this was interesting because it allowed a more conventional interaction style (touching), which could be studied and compared to the other styles.

Fig. 2. The interactive floor about the transportations' culture (top) and the virtual book, which can be browsed by page-flipping gestures performed in mid-air (down)

Next, as illustrated by Figure 2, were another two interactive installations: an interactive floor illustrating the cultural aspects of the means of transportation, e.g. how people gathered together and exchanged goods when a ship arrived to the harbor (in the 1920's this was considered a major event). The design idea was to tailor the interaction style to the content being conveyed, i.e., make visitors walk over in order to get to know the way people walked and moved in the old days.

A second part of the exhibition featured two non-interactive installations: an immersive room featuring a 360-degree projection of traditional fireworks, which is illustrated in Figure 3, and a ceiling projection depicting the process of designing Arab mosques' ceilings.

The design rationale followed for the immersive 360-degree room, was to grab the attention of a second target group of visitors: tourists. The idea was to show visitors a glimpse of Madeira island's famous New Year's Eve fireworks display, a major event that attract thousands of tourists every year, and is currently Guinness World Record for the largest fireworks show on Earth. Since the typical Madeira Islands tourists are over fifty years old, both the artistic and multimedia staff decided not to use another interactive installation, but instead to design for contemplative minds. This ended up as the most enjoyable installation: visitor's reaction was tremendously positive and every visitor, especially children, devoted more time to this installation than to any other. This suggests that the interactivity factor is not necessarily crucial to the success of an installation, if we regard success as merely the visitor's level of satisfaction. We will discuss this a bit more in the next Section of this paper.

4 Discussion

The main customers for this Cultural Interactive Exhibition were architects, artists and designers who were interested in conceiving the best possible ways to provide an interesting exhibition. The design and development processes were therefore a collaborative effort between our interaction design team and these user groups. The interplay between these two groups was very interesting, during the design phase. This comment expressed by one of the designers/programmers was common to many others: "*Change was constant - and communication was a true challenge, since it was difficult for them to communicate us the whole point of. And when we moved from the laptop to an actual kiosk or projection, we noticed how different their opinion was regarding every aspect of the design and development*" (Developer/Designer).

Another interesting observation was: "They [the artists] were completely focused on the MS PowerPoint model - they thought kiosks and interactive installations had to be designed and programmed as if they were PowerPoint presentations" (Developer/Interaction Designer).

Therefore, in a cross-disciplinary project like this, the main challenge was related to communication issues with the artistic direction of the exhibition. Traditionally, artists regard the engineers as mere functionality builders, and the engineers regard the artists as creators of impossible-to-implement products. Adding to the difficulty, there is the language barrier, which is full of jargon (both for artists and for engineers).

The creative use of sensor-based technology was crucial to the success of the project. In practice, how was innovation promoted and how can this be better achieved? First of all, we observed that traditional techniques for fostering creativity still apply. Brainstorming is still very useful, practical and quick. Secondly, we observed the importance of natural interaction: the virtual encyclopedia can be used the same natural way one uses a real book: by simply flipping its pages. The interaction style chosen should be naturally coherent with the content and message conveyed by the product. Speed is always an issue that researchers usually don't have to concern about, but a crucial issue for companies. This implies that whichever innovation process we adapt and follow, it has to be an agile process, given the state of today's turbulent business environment. The ability to make changes quickly is also desirable. In this particular exhibition, the client wasn't happy with the full contents of the interactive installations and continued asking for small additions or corrections almost until the first opening hour of the exhibition.

We organized the remaining most interesting issues around the following themes: usability problems found during visitor's interactions, social interaction issues, differences between the interaction styles, and learning effectiveness.

While some of the installations were created solely as experiential activities, providing an increase in the level of learning by adding facts to an already well-formed body of knowledge, others were designed to support reflective learning [14]. Furthermore, we were also interested in observing the group behavior of the visitors, finding out how collaborative activities can be supported as a feedback mechanism to enhance engagement and learning motivation.

4.1 Usability Problems Found

In general, children had no difficulties interacting with the installations, even though most of them had never used interactive installations of the kind. Older visitors were more averse to interacting and preferred to simply watch. Some visitors didn't immediately realize the installations were interactive: people are not used to this kind of exhibitions. Affordances are needed so that visitors know they can interact with the installations: this would have grabbed much more attention and visitors. For instance, the interactive floor's first version displayed a series of ancient photos and when people walked over it, the photos changed using a mosaic-like effect. After the initial day of observation, we had to add a "walk over me" label, since adult visitors never had a clue and didn't realize that it was supposed for them to walk over the floor. This didn't happen to children, who understood what the floor was for, even with no label or any other clue. Traditional HCI design principles, in this case the importance of affordances, also apply to interactive museum installations. Since these interactions are a relatively short usage time experience, if the visitor doesn't understand the installation's purpose and usage in the initial minutes (even seconds), then the installation won't be effective in supporting the user's visit.

Other usability problems derived from visitors "carrying" the interaction styles from one installation to the other. For instance, after interacting by touch with the historic timeline, visitors moved on to the interactive page-flipping book. Although visitors quickly understood all they needed to do were page-flipping gestures in mid-air (and they quickly understood that the system used motion sensors to interpret the gestures), they also assumed the book's screen was touch-sensitive, and started interactions by touch, which weren't supported by the hardware, of course. The same happened when visitors reached the final two installations, which were not interactive: They assumed they were also interactive, and made pointing and waving gestures. However, this didn't influence the success of these installations, which were perceived as being very engaging by visitors of all ages.

Finally, another usability defect that we discovered only through the questionnaires and surveys was the lack of an index and table of contents to the virtual book. However, we didn't implement this feature, since the screen wasn't touch-sensitive. One advantage of the page-flipping style is that it's too close to the real paper-based book and doesn't provide the navigation means of the digital media, like hyperlinks.

4.2 Social Interaction

As we mentioned, there were much higher levels of social interaction among children than adults. But the most successful ones were clearly the page-flipping gestures performed at the virtual book and the interactive floor, both of which clearly fostered social interaction between visitors. The interactive floor, in particular, allowed collaborations between visitors that wanted to discover the images underneath them and sparked conversations (about culture) between visitors who interacted with the floor at the same time. A third observation regards the immersive 360-degree room. Since this was a dark room with loud fireworks, we were expecting lower levels of social interaction. However, the opposite happened and visitors' social interactions reached the highest level in this room. One explaining factor might be the mere enjoyment that was manifested because of the immersive factor: It's a show of lights and people were more comfortable to make comments to one another while inside that room.

Finally, we studied how collaborative activities can be supported as a feedback mechanism to enhance engagement and learning motivation. In this sense, we observed the interactive floor was particularly effective, since it was frequent to observe the kind of comments like "look how they walked in those old days" between two or more visitors who were experiencing the interactive floor in the third person, i.e. watching others walking over it.

4.3 Differences between Interaction Styles

We observed that there was a tendency to "transport" interaction styles from one installation to the next, i.e. visitors trying to "page-flip" a touch-screen. The collaboration that was allowed by the interactive floor's walking over style clearly enhanced the level of engagement as well as their focus. Regarding the virtual book, it was interesting to note the important role of affordances, which proved adequate to the interaction style: the visitor browses the virtual book the same way he or she browses through a real book. We were expecting more difficulties from visitors, since it's a particularly unusual interaction style. But as we mentioned in the previous section, *page-flipping* and *walking over* were the most preferred interaction styles.

4.4 Learning Effectiveness

While some of the installations contributed to a better learning experience, such as the virtual book and the historic timeline, others seemed better fitted to contribute to a more enjoyable visiting experience. We found it was important to carefully balance these two, so that the whole experience doesn't degrade, i.e. too much learning contents and/or too much fun without learning. Having this in mind, results suggest that the interactive virtual book was the most effective for learning about the exhibition's cultural contents. However, these contents varied somewhat (although they were very similar in terms of apprehension difficulty), and this means that any comparison is difficult to be accurately made. Overall, the significant conclusion was that, above all, we made no harm. Or to put it differently, we tried to make sure that technology would not be a distracting factor from learning.

5 Conclusions

One of the main problems we face today, when trying to provide museum experiences incorporating interactive technologies, is to find out issues that could guide future designs: in this paper, we learned, for instance, the importance of affordances and clues, i.e. the importance of showing museum visitors that they *can* interact with screens and floors, since people aren't still used to this kind of interactive technologies. During the transition to experience design, the HCI community needs to consider many new perspectives on how interactive technologies can bring value to museums and cultural heritage events and exhibitions.

A great difficulty arises when the final product is actually deployed: interactive installations are difficult to prototype and many aspects are actually impossible to model and test by means of early prototypes: sometimes even the color scheme of the projector, as well as surrounding lighting conditions, caused changes in the digital contents. This doesn't happen with, e.g. mobile applications, where the designer has full access to the end product's look & feel anytime and anywhere.

Interactive, real-world exhibitions like this one provide a fertile playground for studying HCI techniques, and for delivering cultural contents to the new "digital natives" as well as to the "digital immigrants" [12].

Some days prior to the exhibition opening, the installations were presented to a group of museum experts, who showed great acceptability towards new digital technologies like these. They were able to explore the interactive installations with care and reported that the interactivity provided an interesting level of engagement. More importantly, several schools visited the exhibition and young students were drawn to the interactive technologies, demonstrating a discernible increase in motivation for learning in the exhibition, as measured by their increased time in the museum and by their informal comments during the interviews.

Despite the challenges and difficulties, we have shown how powerful technology can be, when it comes to digitally augmenting the museum experience.

References

1. Danks, M., Goodchild, M., Rodriguez-Echavarria, K., Arnold, D.B., Griffiths, R.N.: Interactive Storytelling and Gaming Environments for Museums. In: Hui, K.-c., Pan, Z., Chung, R.C.-k., Wang, C.C.L., Jin, X., Göbel, S., Li, E.C.-L. (eds.) EDUTAINMENT 2007. LNCS, vol. 4469, pp. 104–115. Springer, Heidelberg (2007) ISBN 978-3-540-73010-1
2. Hornecker, E., Stifter, M.: Learning from interactive museum installations about interaction design for public settings. In: Kjeldskov, J., Paay, J. (eds.) Proceedings of the 18th Australia Conference on Computer-Human Interaction, OZCHI 2006, vol. 206. ACM, New York (2006)
3. Back, M., Cohen, J., Gold, R., Harrison, S., Minneman, S.: Listen reader: an electronically augmented paper-based book. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 23–29. ACM Press, New York (2001)
4. Garzotto, F., Rizzo, F.: Interaction paradigms in technology-enhanced social spaces: a case study in museums. In: Proc. DPPI 2007, pp. 343–356. ACM Press, New York (2007)
5. Kidd, A.: Technology experiences: What makes them compelling? Hewlett Packard Laboratories, Bristol (2001)
6. Hull, R., Reid, J., Geelhoed, E.: Social impacts of computing: Codes of professional ethics. Pervasive Computing 1(4), 56–61 (2002)
7. SELF-REFERENCE
8. Birchfield, D., Mechtley, B., Hatton, S., Thornburg, H.: Mixed-reality learning in the art museum context. In: Proceedings of the 16th ACM International Conference on Multimedia, MM 2008, pp. 965–968. ACM, New York (2008)
9. Ferris, K., Bannon, L., Ciolfi, L., Gallagher, P., Hall, T., Lennon, M.: Shaping experiences in the hunt museum: a design case study. In: Proceedings of the 5th Conference on Designing interactive Systems: Processes, Practices, Methods, and Techniques, DIS 2004, pp. 205–214. ACM, New York (2004)

10. Mutlu, B., Forlizzi, J., Nourbakhsh, I., Hodgins, J.: The use of abstraction and motion in the design of social interfaces. In: Proceedings of the 6th Conference on Designing Interactive Systems, DIS 2006, University Park, PA, USA, June 26-28, pp. 251–260. ACM, New York (2006)
11. Terrenghi, L., Zimmermann, A.: Tailored audio augmented environments for museums. In: Proceedings of the 9th International Conference on Intelligent User Interfaces, IUI 2004, Funchal, Madeira, Portugal, January 13-16, pp. 334–336. ACM, New York (2004)
12. Prensky, M.: Digital natives, digital immigrants. On the Horizon 9(5) (October 2001)
13. Hall, T., Bannon, L.: Designing ubiquitous computing to enhance children's interaction in museums. In: Proceedings of the 2005 Conference on Interaction Design and Children, IDC 2005, Boulder, Colorado, June 08-10, pp. 62–69. ACM, New York (2005)
14. Norman, D.: Things that make us smart. Perseus Books, USA (1993)

ARAMIS: Toward a Hybrid Approach for Human-Environment Interaction

Stefano Carrino[1], Elena Mugellini[1], Omar Abou Khaled[1], and Rolf Ingold[2]

[1] College of Engineering and Architecture of Fribourg, Switzerland
{Stefano.Carrino,Elena.Mugellini,Omar.AbouKhaled}@Hefr.ch
[2] University of Fribourg, Switzerland
Rolf.Ingold@Unifr.ch

Abstract. In this paper we describe ARAMIS a novel hybrid approach aiming to enhance the human smart-environment interaction. We define this approach as hybrid since it is the combination of three different dichotomies: wearable and pervasive computing paradigms, virtual and real worlds, optical and non-optical sensing technologies. In order to validate the proposed approach we have designed a multimodal framework, in which gestures have been chosen as the main interaction modality. The framework design aims firstly to efficiently manage and merge information from heterogeneous, distributed sensors and secondly to offer a simple tool to connect together such devices. Finally a prototype has been developed in order to test and evaluate the proposed approach.

Keywords: Hybrid gesture recognition, pervasive computing, human-computer interaction, multimodality.

1 Introduction

The interaction between the human and the environment changes along with the technological evolution of society and, with the rise of post-desktop human computer-interaction models such as the pervasive and wearable paradigms, our daily behavior is being changed further. Nowadays specialized research branches are analyzing the interaction between people (Human Centered Interaction) and technological environments [1]. On the one hand lots of works aim at improving the quality of life of inhabitants of smart environments, dealing with the design of new kinds of interactions [2] [3] and trying to enrich the normal way of interacting with the physical environment by technologies seamlessly embedded in everyday objects of our lives [4]. On the other hand the wearable computing paradigm improves the human-computer interaction in a complementary way focusing on: privacy, personalization, and freedom of interaction [5]. In fact, wearable devices can be more intimately and actively involved in the user's activity, and potentially become intelligent personal assistants [6]. Finally with the enhancement of mobile devices performances (processing power, bandwidth and autonomy), it becomes possible to develop new approaches for real-time hybrid systems. In such kind of systems mobile and fixed devices can share information and combine elaboration capabilities and technologies [3].

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 165–174, 2011.

The major contribution presented in this paper is the proposition of a novel hybrid approach, called ARAMIS, to enhance the human interaction with the surrounding environment. Our method is defined hybrid since it is based on three different dichotomies: environmental/wearable paradigms, real/virtual worlds, optical/non-optical sensing technologies. Coupling these technologies ARAMIS aims to exploit the strong points of individual technologies and overcome their weakness (such as light condition in vision-based techniques).

Finally since the gestures are a natural way for a person to interact with the environment and the objects around him [2] [7], we decided, in the context of this research project to consider them as the main human-environment interaction modality.

The paper is organized as follows: Section 2 describes the state of the art of related researches. Section 3 describes and discusses the proposed hybrid approach. Section 4 presents the framework architecture. Section 5 introduces the prototype end the system evaluation. Finally, section 6 concludes the paper and discusses future research plans.

2 Related Projects

Multidisciplinary researches are trying to define what the environment can sense [1] and how it has to react to implicit interactions (with special focus on context awareness) [2] [8] [9]. At the same time, approaches based on on-body, wearable technologies have been widely studied since they can be easily designed to address several typical problems of ubiquitous computing such as: to provide user identification, to supply personal feedback, to improve user activity or user command recognition, etc. [10] [11]. However autonomy, intrusiveness, processing power, ad hoc and not standardized communication protocols still are the major drawbacks of these approaches; even if at present day several works are trying to address these different issues [12].

Both pervasive and wearable computing paradigms have advantages and disadvantages. However several features can be combined in a complementary fashion in order to improve the system performance [5] and the quality of the interaction. According to our knowledge the jointly use of these paradigms, coupled with mixed reality environments and large spectrum of sensing technologies is not largely studied. Moreover the realization of a framework managing this heterogeneous set of sensing data could help researcher in the HCI field to prototype their applications.

3 ARAMIS Hybrid Approach

In order to clarify our model some definitions are needed. In this section we will describe the hybrid approach we proposed, providing definitions and presenting our conceptual structure, stressing the theoretical advantages of this approach.

3.1 Definitions

We have called our approach hybrid for three different reasons:

- A coupling of pervasive and wearable paradigms. In this paper we intend as pervasive paradigm all the information processing (acquisition, transfer, elaboration and presentation) managed at environment level, whereas as wearable the information processing on the human.

- A coupling of real and virtual worlds (object and tools). Real elements could be bound to enhanced digital counterpart. Interactions with virtual objects can change the state of real ones and vice versa. This augmentation of the environment around the user makes possible new kinds of interactions. For instance a user can virtually extend his arms and interact with remote augmented tools.
- A coupling optical and non-optical sensing technologies. Cameras, gyroscopes, accelerometers, and other sensors useful for gesture interaction and context awareness are employed in order to increase the robustness of the dataset available for processing.

3.2 Discussion

Merging together several approaches can allow the system to take advantages of different approaches and technologies. Pervasive and wearable paradigms allow bringing the technology in the environment and on the human itself. A mixed reality environment increases the interaction possibilities beyond the limits of material world and physics rules. Finally combining more acquisition technologies make possible to have a more robust and accurate dataset. So far, several technologies (inertial, acoustic and optical) have been investigated. Those studies showed that each technology has its own advantages and drawbacks and usually it is adapted to specific use cases. For instances optical technologies give very useful information but are expensive in term of processing power and, in addition, sensible to lighting changes; whereas inertial sensors are more resistant to environmental changes but suffer for systematic and incremental errors. Based on these observations, we propose a framework seamlessly integrating several technologies that can be used in a complementary or alternative fashion. For instance we can use both the inertial and optical technology in a redundant manner to provide information about the user gestures or we can use them in a complementary fashion; the elaboration of the images acquired from a camera (expensive in terms of processing requirements) will be processed at environmental side whereas the inertial sensors will be analyzed by the machine on the user.

A strong point that needs to be further stressed in the proposed hybrid approach is how to merge the advantages of the pervasive and the wearable paradigms. The table 1 (an extension of the researches presented in [5]) compares the two approaches resuming their main features.

Table 1. Features provided by Pervasive and Wearable paradigms

FEATURES	*PERVASIVE*	*WEARABLE*
Privacy		**X**
Personalization		**X**
Consistency		**X**
Localized information	**X**	
Localized control	**X**	
Resources	**X**	

Privacy: the target, the medium and the result of an interaction are visible only for the interaction performer.

Personalization: interaction that can be personalized for a user.

With the term *consistency* the prolonged, continuous interaction between the human and the device is taken into account. This exchange of information allows classification algorithms to evolve along with the interaction with the user. In fact an improvement of performances is possible in a twofold manner: the system toward the user and, at the same time, the user toward the system.

Localized information and *localized control*: information and controls are localized in special spot or interactive areas.

A pervasive approach has, in most of the cases, less restriction in terms of *resources* (energy, processing power, etc.).

Some of that features, such as personalization and consistency, can be integrated effortless using the pervasive and the wearable paradigm at the same time. However privacy and personalization in the interaction are further increased by our approach. In fact, information can be fitted and displayed for specified users (according to their profiles and interaction context) exploiting virtual elements.

Another strong point in our hybrid approach is the resource management. In fact, information is elaborated in the mobile system or/and in the fixed one, according to the type of information to handle, the available devices and the quality of the connection. For instance, data that is troublesome to transfer, such as a high frequency data from the accelerometers, can be elaborated locally on the wearable machine and just the processed information (e.g. extracted features) is sent further. On the other hand information that is heavy to elaborate (e.g. images) can be elaborated at environmental side.

The ARAMIS framework, presented in the next section, realized to exploit the advantages of coupling these paradigms and technologies aims to be a useful tool for every researcher that need work in multi-paradigm and multimodal domains.

4 ARAMIS Framework

In order to validate the described approach there is the necessity of a system able to acquire, transfer, elaborate and present the information coming from heterogeneous, distributed (on the users and the environment) sensors and actuators under real time constraints. To address these challenges and to provide a technical support for our theoretical approach we design a framework able to handle these kinds of issues. Designing the framework we chose to manage data as in multimodal systems. This allows exploiting previous knowledge and results from multimodal research area in order to design, prototype and evaluate our approach with widely accepted methods.

Moreover this framework can be a valid aid for researcher working with sensors LAN or WAN in different contexts providing a flexible structure to connect sensors and classification algorithms. In next paragraphs our framework is described taking into account the gesture as main interaction modality, whereas extensions toward other modalities will be briefly discuss.

ARAMIS framework components are presented in figure 1. The Environment Data Module (EDM) and On-Body Data Module (ODM) are the interfaces with the user. They are basically components working as drivers. These modules acquire data from sensors and elaborate them for further transmissions and elaborations. EDM and ODM handle contextual information related to the environment and the user, describing the interactive devices available in the environment, what services they offer and how to interface with them.

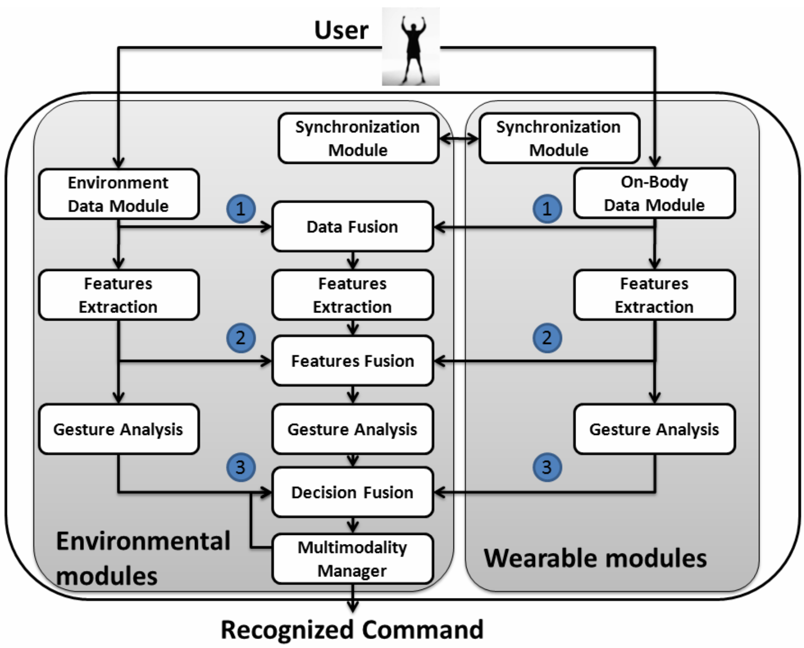


Fig. 1. ARAMIS framework – Overview

According to our hybrid approach the EDM and the ODM represent, respectively, pervasive and wearable components.

As shown in fig. 1, the information originated from each sensor can follow three data flows (numbered in the figure). In the first case, raw data are immediately merged in the *Data Fusion* module (data level fusion [13]). In the second one, information from the environment and the wearable devices is elaborated separately: features are extracted (*Features Extraction* modules) and then merged (feature level fusion). In the last case, separated *Gestures Analysis* modules perform the data classification; hence gestures are recognized separately and finally merged in the *Decision Fusion* module (decision level fusion).

In addition *Data Fusion*, *Features Fusion* and *Decision Fusion* modules have a double goal: firstly to merge the information and secondly to synchronize it. In fact the information is produced from different sources, typically working asynchronously and at different data rates.

The framework allows organizing the information according one of the three configurations for each sensor. Although this allows great flexibility to a system developed using our framework, the drawback is in the realization of the fusion and decision modules. Changes in the system configuration will comport subsequent modifications in these modules or in the use of more complex and often less valid algorithms. At the moment, this selection is done statically by evaluating the available sensors, the target application and the gesture recognition performances needed.

The fig. 1 highlights the case of gesture recognition; nonetheless any modality can be acquired and processed according to the same schema. Hence, even if it is not shown in the figure, the Multimodality Manager module will handle the fusion of the different modalities at decision level.

4.1 Wearable and Pervasive Paradigms as Modalities

From a conceptual perspective, the framework handles sensing data as in a standard multimodal system. This choice grants the possibility to bring the knowledge developed in the multimodal HCI field directly into systems implemented using this framework. For example models as CARE and CASE [13] previously linked to the multimodal research domain can be used to design and characterize systems developed with the ARAMIS framework. The CASE model specifies the multimodal communication types at machine-side. Therefore information from the two paradigms can be used in the system in a concurrent way (e.g. two distinct tasks in parallel, no co-reference), in alternate way (e.g. a task with temporal alternation of modalities), in synergistic way (e.g. a task, in parallel, using several co-referent modalities) or exclusive way (one task after the other using one modality at time, no co-reference) (CASE model). The CARE model specifies system usability properties as complementarity (multiple modalities are to be used within a temporal window to reach a given state), assignment (only one modality can be used to reach a given state), redundancy (multiple modalities, here the pervasive and wearable paradigm, have an equivalent expressive power and are used in the same temporal window) and equivalence (it is necessary and sufficient to use any one of the available modalities). These models have been used to prototype and in the future to evaluate our approach.

5 ARAMIS Prototype and Evaluation

A first version of the framework has been developed in the context of a smart living room. This environment provides rich and interactive features where many devices are able to communicate and interact among each other [14].

Due to the high number of tasks involved in the realization of the whole system (multimodality, gesture recognition, system synchronization, etc.) in the first development phase we focus on demonstrating the feasibility of our approach. Afterward an evaluation of the prototype has been performed in order to assess two main aspects: flexibility of the system, measuring the improvement in the elaboration time using the ARAMIS framework in several configurations, and the changes in recognition rate of gesture commands in different configurations.

5.1 Set-Up and Scenario

As base for the implementation we chose a set-up allowing the realization of a simplified interaction scenario. In this setting we have as pervasive components: an interactive wall (used as augmented interaction feedback), Bluetooth-controlled switches, and a central processing unit performing the data elaboration task. As wearable components the user has: two wireless accelerometers (worn on the arms), a camera (on one wrist) and an Ultra-Mobile PC (at the belt). The final goal is to interact through gestures with a media center device and blinds to adjust the luminosity in the room.

Exploiting data form the on-body module the system is user-location aware: gestures performed in front of the interactive wall's camera are recognized exploiting both the accelerometers and the camera, whereas a reduced gesture vocabulary performed elsewhere is recognized only by the inertial sensors.

Information between the different components is serialized in XML messages in order to include rich metadata (e.g. a time stamp and device ID are added to data gathered in the EDM and the ODM). Hence, a clock synchronization algorithm has been developed to ensure a known global time. The used algorithm (inspired by [15]) guarantees an average error in the clock synchronization below 3ms for our system. It means a clock difference of around 10ms running the synchronization algorithm each 20 minutes (to compensate the CPUs clock shift).

5.2 Environmental Module

At environmental side our gesture vocabulary is focused on dynamic gesture recognition, realized with bare hands, based on computer vision techniques (in front of a camera and facing the wall, fig. 2). The software manages a video stream following these steps: data acquisition, hand segmentation, hand tracking, gesture segmentation and finally hand gesture recognition.

The whole gesture recognition process is realized following the next main steps. Firstly, hands are segmented by skin color. Colors are extracted either statically using the LAB color space (robust against light changes [16]) or dynamically executing a face detection algorithm (based on a Haar cascade classifier [17]) and recognizing nose and cheeks. Then, the two hands and the face are tracked using CAMshift algorithm [18]. Gestures are segmented exploiting kinetic features of hands and finally gestures are recognized through a direction vectors approach. In this prototype we use

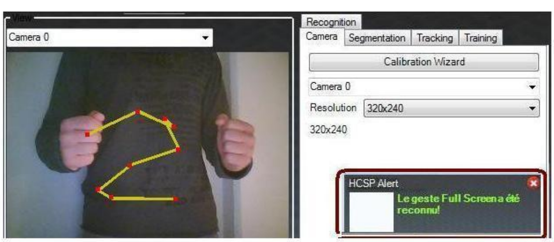


Fig. 2. Environmental module - Hand gesture recognition

a simple gesture vocabulary of height commands: sweep left, right, up, down, both with one hand and two hands). A visual feedback of the hand tracking process is presented to the user.

5.3 Wearable Module

A wearable embedded module for gesture recognition has been developed (fig. 3 shows the sensors on the user arm). Data, received by several sensors (in our set-up inertial sensors and a camera) are sent to an Ultra-Mobile PC. This device works as a bridge with a main server. This module can be configured to work according any of the three paths presented above: as data router (path 1), extracting features from the raw data (path 2), or recognizing gestures and sending the result to the central unit (path 3). Four gestures are recognized (rotations of the wrist, rotations of the wrist two times, shakes of the hand in two directions, pointing gesture). To realize the pointing gesture a customized version PTAMM algorithm (presented by Castel et Al.[19]) has been used. Our implementation allows the tracking in a three dimensional space of the user position related to some interactive objects and tools. Hence this technique has two interesting outcomes: firstly it is possible to make context-aware applications based on the user position (in this scenario the system knows when the user is facing the wall-camera). Secondly it becomes possible to manage the user identification.

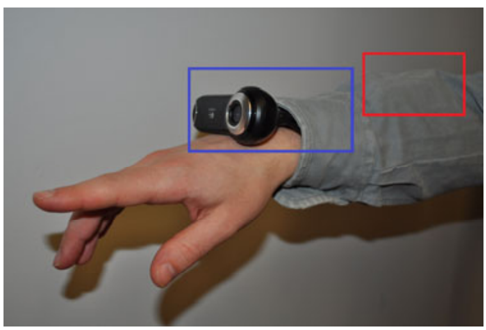


Fig. 3. Wearable sensors. Indigo (left) - camera. Red (right) - inertial sensors.

5.4 Evaluation and Results

In the evaluation phase we have concentrated our efforts to demonstrate the feasibility of the proposed approach deferring usability evaluation to a second test phase. From a technical point of views we evaluate the systems paying attention to real-time issues. We tested our system obtaining elaboration time between 150 and 750ms according to the configuration adopted. The most of this time is spent by the system sending data (serialization, transfer, and deserialization) between devices on the user and in the environment. In the best configuration (150ms between the gesture performance and the feedback) was still perceived as an acceptable response time by the users in our interaction scenario. However, at this moment, the proposed gesture language needs more assessments to be deeply evaluated independently of the worn devices (still too cumbersome).

Differently from the time performances, the gesture recognition rates did not show significant variations according the system configuration adopted. In the context of our set-up we consider these results as not surprising but deeply tests with a more complex gesture vocabulary are planned for the future.

6 Conclusion

In this paper we proposed a novel hybrid approach to enhance the interaction between the human and smart environments. The proposed approach exploits different paradigms and techniques at the same time: pervasive and wearable computing, a mixed reality world to enhance the interaction, and optical and non-optical technologies. In order to validate our approach we designed a framework and we realize a first prototype based on it. Results of the assessments were encouraging: gesture interaction exploiting our hybrid approach can be performed in real time, taking advantage from wearable and pervasive sensors to realize new kind of interactions.

As next steps we plan to perform quantitative measures about system usability and increasing, at the same time, the easiness to connect new sensors to the framework using a service based approach; from that we will complete the formalization of the presented interaction perspective defining a general multimodal language (vocabulary and semantic) suitable for the described approach in a precise scenario.

References

1. Krumm, J., Shafer, S.: How a smart environment can use perception. On Sensing and Perception, 1–5 (2001)
2. Cook, D.J., Das, S.K.: How smart are our environments? An updated look at the state of the art. Pervasive and Mobile Computing 3, 53–73 (2007)
3. Satyanarayanan, M.: Pervasive computing: Vision and challenges. IEEE Personal Communications 8, 10–17 (2001)
4. Cook, D., Das, S.: Smart Environments: Technology, Protocols and Applications. Wiley Series on Parallel and Distributed Computing (2004)
5. Rhodes, B., Minar, N.: Wearable computing meets ubiquitous computing: Reaping the best of both worlds. In: International Symposium on Wearable (1999)
6. Pentland, a.: Smart rooms, smart clothes. In: Proceedings of Pattern Recognition, pp. 949–953 (2002)
7. Bolt, R.a, Herranz, E.: Two-handed gesture in multi-modal natural dialog. In: Proceedings of the 5th Annual ACM Symposium on User Interface Software and Technology - UIST 1992, pp. 7–14 (1992)
8. Schmidt, A.: Implicit human computer interaction through context. Personal Technologies 4, 191–199 (2000)
9. Schmidt, a.: There is more to context than location. Computers 23, 893–901 (1999)
10. Culler, D.: TinyOS: Operating System Design for Wireless Sensor Networks. Sensors (2006)
11. Nixon, P., Wagealla, W., English, C., Terzis, S.: Security, privacy and trust issues in smart environments. In: Smart Environments: Technologies, Protocols and Applications. Wiley-Interscience, Hoboken (2005)

12. Argyroudis, P.G., O'Mahony, D.: Securing communications in the smart home. Embedded and Ubiquitous Computing, 133–139 (2004)
13. Dumas, B., Lalanne, D.: Multimodal interfaces: a survey of principles, models and frameworks. Human Machine Interaction (2009)
14. Mugellini, E., Khaled, O.A., Pierroz, S.: Generic Framework for Transforming Everyday Objects into Interactive Surfaces. Human-Computer (2009)
15. Römer, K.: Time synchronization in ad hoc networks. In: Symposium on Mobile Ad Hoc Networking and Computing (2001)
16. Cai, J.: Detecting human faces in color images. Image and Vision Computing 18, 63–75 (1999)
17. Papageorgiou, C.P., Oren, M.: A general framework for object detection. In: Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271), pp. 555–562 (1998)
18. Bradski, G.: Computer vision face tracking for use in a perceptual user interface. Interface (1998)
19. Castle, R.O., Klein, G., Murray, D.W.: Combining monoSLAM with object recognition for scene augmentation using a wearable camera. Image and Vision Computing 28, 1548–1556 (2010)

Express Yourself: Designing Interactive Products with Implicitness to Improve Social Interaction

Huang-Ming Chang[1] and Rung-Huei Liang[2]

[1] Eindhoven University of Technology, Department of Industrial Design, P.O. Box 513, 5600 MB Eindhoven, Netherlands
[2] National Taiwan University of Science and Technology, Department of Industrial and Commercial Design, No.43, Sec. 4, Keelung Rd., Da'an Dist., Taipei City 106, Taiwan

Abstract. In this paper, we explore the possibility of designing interactive products that allow users to express themselves with daily behaviors to have their own special experiences, and the expression would improve their social interaction as well. Three design works were implemented. For each design, experiment was held to collect user feedback, which was mostly positive. We also see a tendency that users would be much expressive when interacting with physical objects that bear more information related to their everyday life.

Keywords: Expression, Interactive Products, Implicitness, Social Interaction.

1 Introduction

According to the vision of Ambient Intelligence, the role of technology should be shifted from use to presence[1]. A 'use' centered view focuses on functional aspects of a system or device. Viewing a device or system from a 'presence' perspective entitles users to get more involved in the interaction. Ross and Keyson[2] also argue that design should evolve from task-oriented to experience-driven one. By considering the role of a user, we propose that one could be invited as a co-author of the experience while interacting with interactive products. Therefore, users are more regarded as participants rather than audiences who passively appreciate the work created solely by designers. In this paper, we explore the possibility of designing interactive products that allow users to express themselves to complete their own experiences, and others' expression would also induce users to interact with each other. We implement three works that comply with this concept and investigate user experience after use while putting into a community context.

2 Related Work

Most people experience the frustration of being awakened by alarm clocks. However, the alarm clock could never perceive the negative emotion expressed by the user. To tackle this one-way communication, Wensveen, Overbeeke, and Djajadiningrat[3]

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 175–184, 2011.

design an alarm clock that allow users to express their feelings which could also be recognize by the clock, and the clock would adjust its reaction, playing more proper music (see Fig. 1 left). This work indicates the importance of users' expression within an emotionally rich interaction process. It also unveils the opportunity for designers to deal with mundane behavior that might elicit rich interaction by creating space for users to express themselves.

Dance Rail [4] is another related work about user's expression, which is an interactive installation that allow one or more audiences participating in the creation of the whole experience. Audiences could slide their hands along with the touch-sensitive lighting rail to perform fluent motion, which is made by collecting real body expression performed by a professional choreographer (see Fig. 1 right). Once an audience stepped onto the stage, his or her role would turn into not only a performer but also a participant of the total experience for the exhibition. This work itself is already an aesthetic installation in public space, but it makes more sense if one or more audiences are willing to perform their own expression.

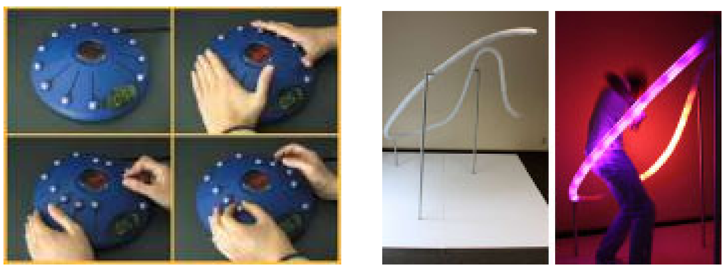


Fig. 1. Wensveen, Overbeeke, and Djajadiningrat[3] design the alarm clock that allows users to express their mood (left). Dance Rail[4] is an interactive installation that invites audiences to express in their own way (right).

The above two works help discover the potential of designing interactive products that encourage users to express their feelings. However, to create the space for users' expression within an experience, the challenge to designers is to trace back the essential of user experience instead of insert several extra steps while using a product. To elicit users' expression, the whole context should be taken into concern, providing efficient emotional motives to induce people to express naturally rather than force them to express as a functional task.

How to bridge the gap between people who are geographically separated has been discussed for years[5-8]. Social media utilize web-based technology to build up dialogue in various ways, such as instant messenger and blogs. However, while social media is pouring all the information to users, one essential of social interaction has been missed. Sociality is often a more subtle and delicate thing[6]. Among close friends or lovers, emotions still could be delivered even if very little information is exchanged. Most of the functional-oriented services provided by social media lead to explicit and direct conversations but ignore the possibility to deliver emotion in an implicit way. Moreover, people tend to utilize spatial arrangement to help themselves keep things in minds [9], which implies that users could retrieve abundant memories

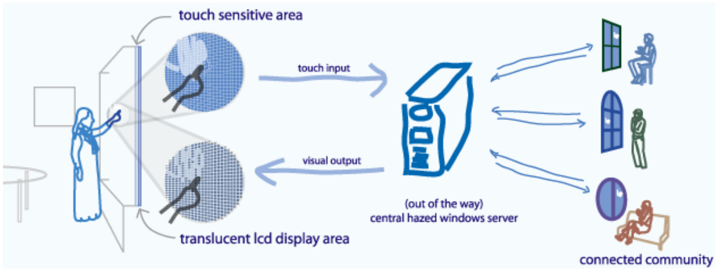


Fig. 2. System diagram of Hazed Windows concept[10]

while interacting with their surrounding environment. Therefore, here comes the challenge to designers: how to facilitate users to express via Internet without disconnected with their real world surroundings?

Hazed Windows concept[10] demonstrates a possible solution to this question, which intended to emphasize the presence of people via Internet. The concept uses everyday experience as a metaphor, drawing in the mist of warm breath on a window. You could make marks on the digital surface that connects to other devices of your loved one. Your marks show up on the other surface as well but fade away gradually over time (Fig. 2). According to Löwgren[11], this work is categorized to peripheral interaction and calm technology[12]. The interaction does not require being at the center of user's attention. Imagine that you might head up and look at the Hazed Window incidentally, and then you see your grand child is drawing something for you. What makes more sense is the presence at that moment rather than the exact information it is delivering. With the intention to deliver the presence and awareness rather than the digital content, Hazed Windows concept rejects the default assumption that digital mediated communication is persistent and reflects the nature of social interaction: only if you notice it, the message remains in your memories. This also demonstrates that the interaction throughout Internet could be implemented in an ambient way without dealing with a typical user interface. Based on the argument of Hallnäs and Redström[1], the perspective of these works focuses more on presence rather than useful functions. Moreover, this work reveals another notion that needs further discussion: could Internet help improve social interaction in real life?

Most of current social media aim to have users connected all the time. Therefore, the function of the system would grow more complete and stronger to fulfill all the needs for users to interact with others. However, all the interactions that occur within

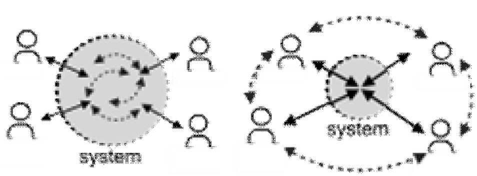


Fig. 3. In previous time, users usually interact with each other within the system (left). This paper intends to evoke social interaction outside the system (right).

the system somehow might reduce the physical social interaction among users. Moreover, current social network services are all web pages or widget-based, users' behaviors could not escape from screens and buttons. The aim of this paper is to simplify the interaction between users within the system and encourage users to have more physical social interaction in their daily life (Fig. 3).

3 Implicitness

To achieve this goal, one essential element is to leave proper space and freedom within the interaction for users to express themselves, which we call "implicitness." Since the social community is not entirely predetermined, developers should leave things incomplete and give opportunities for users to finish the design themselves while designing social platforms[13]. What do "space" and "freedom" mean? Space of expression represents the given opportunities to express. Users should be entitled to express their emotion, such as the frustration when being woken up, the excitement when hearing favorite song, or anything they want to say but no one is around. Freedom of expression implies that designers should not stereotype users, assuming that users have certain behaviors. For example, the title column of e-mail sometimes bothers users because the content is too rough to put a proper title for it. From the perspective of presence, what really matters is to make expression rather than the content of the expression. The challenge for designers is to discover the moment when people would have emotional reaction while interacting with products and to create a proper channel for them to express.

If users have the space and freedom to express, the amount of expression from users would certainly be great. Therefore, "implicitness" also implies that the way to convey users' expression via interactive products should be in an ambient manner to avoid being flooded. Users could be aware of other members' presence in his social network without disturbance, and are always entitled to decide if he or she wants to engage with the social network. However, current social network services tend to be the center of users' attention while interacting with it. A new way of interacting with the social network should be developed, which works only in the background of daily environment.

The most challenging part is to encourage users to interact physically more, rather than totally digitally. "Implicitness" represents the detail of users' expression should be partly hidden, which would give strong emotional motives for users to find the detail if they feel strongly connected with this community. Since the function of the system is deliberately reduced only for expression and awareness, users would seek other ways to discuss with others, such as making phone calls, talking face-to-face, etc. On the other hand, every user would have their personal interpretation of the anonymous expression, which might recall memories or previous experiences.

Therefore, three general principles are made: (1) the process of the interaction should give appropriate space and freedom for users to express themselves; (2) the expression and awareness should be delivered in an ambient manner, being integrated into ordinary daily behavior; (3) the detail of the expression or awareness should be partly hidden to evoke extra social interaction.

4 Research through Design Approach

In order to explore the above principles, we implemented three design works and held experiments to validate our concepts. We picked three moments that users might have emotional expression: listening to radio, being woken up by alarm clock, and finding things in a drawer. First, by interacting with workable prototypes, users have the space to express themselves. The way of making expression is to share music or to record any kind of sounds, including voices. Users are free to express with any kind of auditory information. Second, the delivery of the expression is integrated into everyday behavior peripherally. Users could choose to ignore the incoming message and move on what they are doing, or listen to it to see if anything comes to minds. For the last principle, all the expressions are delivered anonymously.

The form of the product is also another issue that we intend to discuss. Although all the information technically transmits through Internet, if the form of the interface becomes more like daily object, would a user feel more humanity rather than technology? Therefore, we start from fully digital form (website), hybrid form (physical object with GUI), and then physical form (physical object without GUI). With these three working prototypes, we invite three groups of people who know each other as subjects to participate in our experiments and investigate user experiences while interacting with these three prototypes.

The first design, Social Radio, is an online radio website that is categorized into fully digital form. Different from other online radio services that usually have DJs to host programs, Social Radio invites users to upload the music that they would like to share. Second, Social Clock is a physical alarm clock that includes touch screen GUI, which belongs to the hybrid form category. When a user intends to set up the alarm, he or she needs to record a few seconds of sound to be uploaded automatically. All the sounds contributed by users would be randomly assigned to the next alarm sounds of these members' Social Clocks. The last one, Whisper, is made of fully physical form without GUI, which allows one to record a secret and bury it in the drawer. Secrets in the Whisper drawer would be automatically synchronized. One might find an anonymous user's secret accidentally in one's Whisper drawer.

4.1 Social Radio

People like to request songs with special meanings to express their current feelings. Furthermore, listening to radio is usually put aside from the center of user's attention while working. However, some specific music that bears special meanings to users would attract immediate attention because of evoked memory or experience. Therefore, this work utilizes radio as a platform that invites users to share music or recorded sounds and voices, which transforms audiences into expression makers.

A within-subjects experiment was held. Ten subjects familiar with each other were recruited for the experiment and then separated randomly into two groups. One group was assigned to use Social Radio one hour a day while the other group was listening to traditional Internet radio. After a week, two groups exchanged their tasks. After the tasks were performed for two weeks, the participants were instructed to fill out a questionnaire on their feelings and satisfaction level. After that, subjects were interviewed to retrieve user experiences.

Table 1. Results of one-way ANOVA of the feedback of Internet radio and Social Radio Note: N=10, * means P<0.05, significant difference exists

Level (1Minimum – 7 Maximum)	Internet Radio Mean (SD)	Social Radio Mean (SD)	F value	P value
Preference / Acceptance	4.8 (1.87)	5.3 (1.06)	0.178	0.472
Interesting	3.7 (1.77)	5.8 (0.92)	0.308	0.004*
Expecting	4.7 (1.77)	5.6 (1.07)	0.474	0.186
Knowing about other members	3.5 (1.06)	5.7 (1.72)	0.129	0.003*
Preference of the interactive manner	3.6 (1.26)	5.4 (1.17)	1	0.004*

Results of one-way ANOVA indicate the difference between the experiences of Internet radio and Social Radio. The two groups significantly differed (p value < 0.05) in the degree of interest, knowing of other members, and the preference of the interactive manner (Table 1). Therefore, the two groups were compared in terms of the mean value. The feedback of Social Radio group with respect to feelings after listening was more positive than that for the control group. Furthermore, Social Radio listeners were more aware of the recent circumstances of other members than Internet Radio listeners were. From the interview, subjects mentioned that Social Radio basically is Internet radio to them, which makes no significant differences, but the sharing scheme within the group is interesting. They could share their own favorites and could almost guess the contributor of the song that is playing on Social Radio.

4.2 Social Clock

Being woken up usually makes people feel negative emotion: you know you are really sleepy but still have to get up when the alarm sounds. Some people prefer to be woken up in a more humane way, such as phone calls from friends, which would make them feel better. On the other hand, while being woken up by the alarm clock that the users set by themselves, the conflict in their minds would result in emotional reaction, which would be a proper opportunity for them to make expression. With Social Clock, one could record voice by Social Clock anytime, which would be a clip of alarm sound of one's friends. In the next morning, a user would be woken up by one of his friends' voice, which is selected randomly and anonymously. Although the general function remains the same, the role of the clock is slightly transformed into a stage that allows users to make expression.

Fig. 4. Social Clock. Subjects participated in the test of simulated woken-up experience of traditional alarm clocks in comparison with Social Clock.

This experiment applies between-subjects to determine if the Social Clock significantly improves the experience of being woken up. Twenty-four volunteers, including sixteen people who know each other (in-group subjects), were recruited for this experiment and separated into control group and Social Radio group. To simulate the scenario of being woken up, a comfortable bedroom space was set up for each participant to enter individually. Participants were relaxed with a soft light, calm music, and a fragrance oil to induce sleep. Each subject was led to the bed, instructed to set up the alarm clock, and then to take a nap for one hour until the alarm sounded (Fig. 4). After being woken up, the participants were instructed to fill in the questionnaires and were also interviewed regarding their experience of being woken up.

Table 2. Results of one-way ANOVA of the feedback of traditional clock and Social Clock Note: N = 24, * means P<0.05, significant difference exists

Level (1Minimum – 7 Maximum)	Traditional Clock Mean (SD)	Social Clock Mean (SD)	F value	P value
Preference / Acceptance	3.00 (1.41)	5.38 (1.06)	0.742	0.002*
Interesting	1.75 (1.04)	5.88 (0.64)	0.281	0.001*
Expecting	1.88 (1.13)	5.63 (1.19)	0.761	0.001*
Knowing about other members	1.88 (0.99)	5.75 (0.89)	0.924	0.001*
Preference of the interactive manner	2.13 (0.83)	5.50 (1.20)	0.270	0.000*
Have better mood	1.63 (0.52)	5.50 (1.20)	0.016	0.001*
Family, close friends	2.13 (0.83)	5.75 (1.04)	0.550	0.001*
Inviting people to join	1.88 (0.64)	5.38 (1.19)	0.558	0.001*

Social Clock brings positive experience of being woken up in every aspect listed in Table 2. According to the interviews with subjects, the main reason leads to this result is the space for expression and the unpredictable music or sounds. Non-group subjects tended to share general pop music. On the other hand, those in-group subjects recorded private messages, funny sounds, or music with special meanings. "I want to collect appropriate music for each of my close friends," said by subject #14. "I want to share some music about current social events or recent festivals," said by subject #04. Subject #20 even shared a long narration to the listener that she imagined, "Hey, wake up, make yourself a cup of coffee, …the girl you love is waiting for you. Don't forget to bring her a bunch of flowers."

After being woken up, in-group subjects tend to make a guess of the identity of the contributor. "This definitely is shared by her! Only she would listen to this song!" said by subject #07. Subject #01 said, "I want to take a revenge!" after being woken up by funny sounds. Nonetheless, some non-group subjects concern about if some unknown user would share scary music. "Can I filter out those songs not suitable for waking up?" said by subject #08. It seems that differences exist between in-group subjects and non-group subjects. These emotional reactions convey the intention to start social interaction afterward, which complies with what we claim above.

4.3 Whisper

This work removed all the typical switches from the interaction between users and products that deliver users' expression via Internet. Furthermore, the challenge is to integrate the process of making expression into daily behavior seamlessly. From observation of daily behavior, we noticed that people tend to put their tiny, private things in their drawers, which is also a moment full of emotions. Moreover, the behavior of putting things inside and finding things out is similar to that of burying and digging in the ground, which would be a proper metaphor to integrate the opportunity of expression into daily behaviors. Several LED lights were embedded under the acrylic board at the bottom of the drawer to represent the buried secrets. When one tells a secret to the empty space in the drawer, it will record the secret. On the other hand, if one accidentally discovered the light flashing at the bottom of the drawer, the secret would play automatically and then disappear, like mysterious whispers from the land.

Fig. 5. Subjects discover other's secrets and burry their own secrets in the drawer during the experiment (left). The space for experiment was set up as a normal study room (right). Working prototype of Whisper (highlighted) is attached to a small cabinet.

Table 3. Results of descriptive statistics of Whisper

Level (1Minimum – 7 Maximum)	Max	Min	Mean	SD
It is an intriguing product	7	4	5.95	0.826
Want to hear others' secret	7	4	5.60	0.995
Able to guess who share the secret	7	3	5.55	1.191
Tend to imagine the stranger's life	7	3	5.70	1.129
Willing to share my secret	7	3	5.25	1.118
Curious about who took my secret	7	1	5.60	1.465
Want to share the secret just heard to your friends	7	2	5.10	1.483
Looking forward to hearing more secrets	7	4	5.50	1.147
Willing to have one if my friends all have one	7	3	5.70	1.174
Inviting friends to join	7	3	5.60	1.046

Twenty volunteers were recruited to participate in the experiment (Fig. 5). Since there is no similar product to compare with, this experiment focused on understanding users' feedback. For each session, one subject was asked to share what they do for social interaction via Internet, and then started to use Whisper. After using Whisper, the subject was instructed to fill out the questionnaire about Whisper (Table 3). The descriptive statistics of the questionnaire generally shows positive results (lowest is 5.10/7.00, SD=1.483). Besides, subjects showed their creativity on making expressions that reflect their moods or events in their life. Subject #03 said thanks to the stranger who helped her. Subject #07 said to her cat, ”Cute Kitty don’t be sad. I am coming home next week!” Subject #02 said to his ex-girlfriend “Happy Birthday to you though you might not be able to hear this. I still love you.” Subject #06 apologized to his friend for long ago “I knew you were mad, so that I didn’t dare to admit it was me.” Subject #12 said, “Actually, I am getting baldheaded.” Subject #19 didn’t say anything, but he sang a song. Subject #04 shared funny things about subject #10. Subject #10 coincidently found it and responded, “I’m going to kick you!”

5 Discussion

Since these three works were based on different contexts, the results could not be compared directly. However, we see a tendency that subjects were most willing to express via Whisper, second is Social Clock, and the least is Social Radio. The cause is complex, but we tried to generalize possible explanations of it.

The form of these three works might affect the result. Subjects needed inspirations to make their expression. However, inspirations mostly originate from their surroundings, which means that the amount of expression might depend on the variety of things related to subjects’ life. For these three works, forms are categorized into three types: digital, hybrid, and physical. We would argue that subject would encounter more inspiration while interacting with more physical objects. For example, while subjects interacting with the GUI of Social Radio and Social Clock, subjects of Whisper is touching and seeing physical objects in the drawer, which might help retrieve more memories in life.

Subjects’ feedback collected from interviews, though no direct evidence that supports to improve physical social interaction, reveals the intention of the subjects to start new conversation after hearing the expressions that others made, such as “This definitely is shared by her!“ “I want to take a revenge!” and “I’m going to kick you!”

6 Conclusion

We explore the possibility of designing interactive products that let users to express themselves with daily behaviors to make their own special experiences., which also encourage them to interact with each other in their daily life. We also found that users would be much expressive when interacting with physical objects. Therefore, we suggest designers to utilize physical form while designing computational products with this concept. With this research, we see the advantages brought to user experience through inviting users to express themselves within the interaction process, which also reveals more potential on the direction of transforming users from passive to active while interacting with products and space in response to the up-coming age of the Internet of Things, when objects are capable of sensing, responding, and transmitting information.

References

1. Hallnäs, L., Redström, J.: From use to presence: on the expressions and aesthetics of everyday computational things. ACM Transactions on Computer-Human Interaction (TOCHI) 9, 106–124 (2002)
2. Ross, P., Keyson, D.: The case of sculpting atmospheres: towards design principles for expressive tangible interaction in control of ambient systems. Personal and Ubiquitous Computing 11, 69–79 (2007)
3. Wensveen, S., Overbeeke, K., Djajadiningrat, T.: Push me, shove me and I show you how you feel: recognising mood from emotionally rich interaction. In: Proceedings of the 4th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques, pp. 335–340 (2002)
4. Toering, E., Man, P., Jong, F.d.: Dance rail: an interactive installation that provokes aesthetic movement. In: Proceedings of 5th International Conference on Design and Semantics of Form and Movement, pp. 206–207 (2009)
5. Chen, C., Forlizzi, J., Jennings, P.: ComSlipper: an expressive design to support awareness and availability. In: CHI 2006 Extended Abstracts on Human Factors in Computing Systems, pp. 369–374 (2006)
6. Strong, R., Gaver, B.: Feather, scent and shaker: supporting simple intimacy. In: Proceedings of CSCW 1996, pp. 29–30 (1996)
7. Dey, A., de Guzman, E.: From awareness to connectedness: the design and deployment of presence displays. In: Proceedings of the SIGCHI Conference on Human factors in Computing Systems, pp. 899–908 (2006)
8. Lottridge, D., Masson, N., Mackay, W.: Sharing empty moments: design for remote couples. In: Proceedings of the 27th International Conference on Human Factors in Computing Systems, pp. 2329–2338 (2009)
9. Kirsh, D.: The intelligent use of space. Artificial Intelligence 73, 31–68 (1995)
10. Freiesleben, T., Knapek, M., Moberg, H.: Hazed windows (2001), `http://webzone.k3.mah.se/kid01016/projekt_forskningstema.htm` (retrieved December 22, 2010)
11. Lowgren, J.: Fluency as an Experiential Quality in Augmented Spaces. International Journal of Design 1, 1–10 (2007)
12. Weiser, M., Brown, J.: Designing calm technology. PowerGrid Journal (1996)
13. Crumlish, C., Malone, E.: Designing Social Interfaces: Principles, Patterns, and Practices for Improving the User Experience 17 (2009)

Mojo iCuisine: The Design and Implementation of an Interactive Restaurant Tabletop Menu

Ting-Han Chen, Hsin-Hou Lin, and Yi-Di Yen

XXtraLab DESIGN Co., 1F, No.33, Sec.2, Zhongxiao East Road, Zhongzheng District, Taipei, Taiwan
{daniel,jimmy_iw,tiffany}@xxtralab.tw

Abstract. This paper discusses the design and implementation of an interactive tabletop system for moJo iCuisine, a western restaurant in Taiwan. The restaurant has 22 interactive tables that enable diners to browse the menu, order dishes, play games, fill out opinion forms, and check their bill. The design and implementation process and findings are discussed.

Keywords: Interactive restaurant, interactive menu, tabletop, service design, ordering system.

1 Introduction

Ubiquitous Computing [1] is shaping the future of computers by integrating computers into everyday objects so that users are able to interact with them intuitively to obtain information or digital services. Employing the concept of Ubiquitous Computing, research projects such as SmartSkin [2] and Everywhere Displays Projector [3] have successfully provided technological frameworks and feasible methods for practical use of Ubiquitous Computing, which transforms spatial elements such as tabletops, floors, or walls into interactive surfaces.

The emergence of this technology, generally referred to as “interactive surfaces” [2, 4], has made it possible for information systems to be embedded into furniture, interior periphery, or everyday objects, and be offered to people in the required context. The space becomes the interface, and relationships among the user, space, and computational service are changed.

Traditional architectural and interior design methods have not yet been updated to take into account this conceptual evolution so that they can facilitate the implementation of interactive surfaces into new spatial designs. The design of interactive surfaces in different spatial contexts is thus become one of the main research issues for the design of novel interactive systems [16].

1.1 Bringing an Interactive Surface into a Restaurant Space

The introduction of interactive surface into specific spaces, such as museums [5], classrooms [6], or collaborative environments [7], has introduced new ideas about

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 185–194, 2011.

spatial design and opened up new solutions to existing problems. On the other hand, commercial spaces, such as restaurants, have business purpose, service model, and system integration issues that require different design goals. This makes them challenging spaces to design for.

Nevertheless, the restaurant industry has been adopting information technology to facilitate restaurant operations and services. This technology has been used in areas such as accounting systems, POS systems, and online reservation services. Not only backend systems, but also frontend interfaces such as self-service electronic ordering systems [8], have been introduced into restaurants to enhance operations. The original purpose of such ordering systems was to improve efficiency and accuracy. However, they have also created interface design problems since they require user-system interaction and thus require advanced spatial interaction design knowledge for development. Moreover, to improve service quality and to explore operational possibilities, new concepts, new approaches, and new media for restaurant design have also been attempted in the commercial sector [10, 11, 12].

1.2 Design Problems

The aforementioned conditions illustrate the digitizing trend in the catering industry and the ongoing explorative use of interactive systems designed to offer better service. However, there is no clear method for designing an interactive surface (e.g. tabletop) for a restaurant to facilitate service that is improves the experience of both the consumer, employee, and owner. With this in mind, the goal of this study was to design and implement an interactive restaurant that addresses the following questions: How can an interactive tabletop facilitate restaurant operations? How can it be integrated into the interior design and restaurant service system? How should such a tabletop interface be designed? What will the service flow be?

This paper (1) analyzes the functioning flow of a commercial restaurant, (2) examines the design process and implementation through the actual construction and implementation of an interactive restaurant tabletop, mojo iCuisine, where each table in a restaurant was fit with a built-in interactive tabletop, (3) explores inspiration it might provide for, and implications it might have on, spatial interaction design based on phenomena that arose as a result of the design and implementation of this technology, and (4) discusses informal studies on preliminary results for further investigation.

This project focuses on the intersection of interactive surfaces, restaurant service systems, and spatial design. Catering service factors, including internal quality considerations such as dietetic hygiene, servers' attitudes, and different dining cultures, are not within the scope of this paper. The focus of this study is the issues that emerged as result of the design and implementation of an interactive tabletop in a restaurant, how they can help spatial interaction designers, and the possiblilty of developing other value-added services.

2 Relevant Works

Relevant technical research into interactive surfaces [2, 3, 4] was studied help us practically implement the interactive tabletop system. Previous studies on restaurant interactive systems and spatial design concepts and strategies were also reviewed to glean insights into, and directions for, the design of a restaurant incorporating an interactive surface system.

The idea of interactive systems for restaurants is nothing new. There are many examples. The integrated interactive restaurant communication system [8] patented by Kurland and Gilbert in 1985 is a relatively comprehensive early case. Diners were able to place their orders directly using a television on the table this bypassing waiters. While waiting for their meals they could play games on the television.

More recent studies, such as MenuVista [9], Inamo [10], uWink [11], and e-Menu [12], mostly discussed the use of interactive tabletops to allow diners to place direct orders and speed up meal service. uWink in Los Angeles bills itself as an entertainment restaurant where diners can play electronic games on the built-in touchscreen on the table while dining. e-Menu, designed by Conceptic, is a touchscreen interactive menu hung on the side of a table. Inamo in London uses projectors to dye the menu on the table, which as a touch-tabletop, allowing diners to select items using a pointer controlled via a small square of touchpad embedded into the tabletop. While MenuVista also paints the image on tabletop using a projector, its user input is based on capacitive sensing where the capacitive sensing screen was inlaid under the tabletop preventing the sensitive electronic screen from getting wet or dirty.

Similar to Inamo, uWink, and e-Menu, mojo iCuisine is an example of an actual business application. However, this paper discusses the design and implementation of an interactive menu tabletop actually implemented in a commercial restaurant.

This paper does not simply examine the design of a standalone tabletop. It specifically examines the design of a tabletop for a restaurant space. Its context and constraints should be considered. As a result, relevant spatial interaction design concepts in the realm of Human Computer Interaction, such as Ubiquitous Computing [1], Ambient Displays [13], Roomware [14], and Interactive Spaces [15] are drawn on to help validate the integrity of the interactive tabletop and ensure the usability of the system, which was designed for a restaurant space using a user-centered design approach.

3 Design and Implementation

3.1 Design Method and Steps

This design project is explorative and service-oriented. Observational methods are commonly used to discover insights as the first step in tackling explorative and service-oriented problems, in both interactive system and service design methodologies. Hence, we started by observing user (i.e. restaurant diners) behavior from a user-centered design perspective and conducted a needs analysis of restaurant owners. We were then able to formulate a design framework as the basis for the spatial interaction design of an interactive restaurant in the second step. Based on the framework laid, and combined with relevant design principles found in previous studies, the interactive restaurant was then designed and constructed. Last, the restaurant began commercial operation using the interactive tabletops to validate the feasibility of the design.

3.2 Formulation of the Design Framework

First, we conducted a needs analysis by interviewing restaurant owners. It was found that owners did not want labor be replaced by technology, but rather to facilitate it, because the intimacy of the interaction between servers and consumers is irreplaceable. Because of this, they want a system that will facilitate the service process and enhance the experience of diners. Finding a way to make an impressive or unforgettable dining experience without detracting form the importance of the meal was also very important. In addition, accuracy and efficiency of information delivery, for example, what has been ordered by which table, was necessary.

Following these requirements and observations of the procedures of similar restaurants, the dining process was divided into six main steps beginning at the time diners are seated and ending at the time they pay the check. The steps are as follow: 1) browsing the menu, 2) ordering, 3) waiting to be served, 4) eating, 5) filling out the questionnaire (if there was one), and 6) paying the check. It was also found that, if not alone, some diners continued to interact (e.g. chat) throughout the process except for when they were interrupted by the service process or their cell phone. Although diners occasionally left their seat to use the bathroom or talk on their cell phone, most of the time diners remained seated interacting while dining process occurred over the tabletop. We were made aware of many design possibilities that could be attempted in the future by examining diners' behavior. This inspired us to establish a modularized system architecture in which service items could be added or modified later.

From the above analyses a basic service model for system was created. It is as follows:

- *Browse the digitized menu and order.* Diners use the system to select set meals, individual items, and drinks. Once the final order is confirmed, it is sent to the kitchen via a Local Area Network, where the printer will print it out and the cooks prepare the food accordingly.
- *Play games while gaiting for the meal to be served.* While waiting for their meals, light games, such as chasing butterflies or rolling the dice, are provided by the system to diners to kill time or catalyze social interaction.
- *Change the digital tablecloth.* Diners can touch the tabletop to select different tablecloth themes, such as night sky, magical stars, or dazzling colors, to alter the dining atmosphere.
- *Check/Pay the bill.* Diners can check their order or as ask for the bill at any time. If diners choose to pay the bill, the system will notify the counter immediately and signal the waiter to collect their money.
- *Fill in the questionnaire.* Upon confirmation that the bill has been paid, the system will load questions pre-set by the management from a database. The diners can answer the multiple-choice questions using the touch-tabletop. Feedback and comments are collected and calculated.

3.3 Physical Structural Design

When designing an interactive surface for a specific space, factors such as architectural characteristics, physical constraints, technical stability, and user-centered design propositions should be taken into account. We utilize the tabletop as the interface and human touch as the direct and intuitive method for diner-system interaction. Based on a spatial usability and ergonomics, the physical structure of the interactive tabletop unit was thus designed (Figure 1).

In this design each table is set with two seats, and equipped with a PC, a projector, and a capacitive sensor board embedded beneath the tabletop. Wires connecting each component are embedded into the architectural space or peripheral artifacts, and hidden from users. As a result the tabletop is identical to that of an ordinary dining table. The PC and projector are hidden into a customized ball-shaped cover, which is integrated into the overall spatial design, hovering above the table like a luminaire setting. Images are projected on to table surface by the projector in the above.

3.4 Tabletop Interface Design

The projection occupies the entire tabletop and the frame of the screen is invisible so that the diners will feel like they are interacting with the space rather than a screen or an electronic device. Since there is one seat on each side of the table there are two interfaces facing opposite with shared features (see Figure 2).

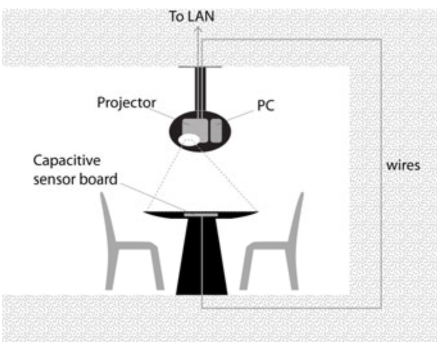


Fig. 1. Physical structural design of an interactive tabletop unit

Fig. 2. Interactive menu interfaces with shared features facing opposite directions

There is a touch-sensing set on each side. Each has four options presented as iconic buttons for personal control. The options include: browse menu and order, change tablecloth, games, and bill. A two-phase checking feature double-checks diners' orders and bill requests. This also prevents unexpected orders resulting from accidental interface commands. There is a rotatable menu in the middle of the tabletop between the personal control interfaces so that when one diner touches the tabletop the same menu information will appear on both sides. This was done to preserve the fun of looking at a menu together.

The layout ensures that the dining area and touch-sensing area are separated to prevent interference. In addition, there is also a touch-sensing lock in the center of the table that allows diners to lock and unlock the touch-sensing feature at any time. This prevents accidental interface commands.

3.5 Backstage Control Mechanism

For full exploitation of the interactive tabletop interface, the implementation of backstage control is very important. We designed several control units and placed them at different places according to their features. Figure 3 below briefly illustrates the overall operation mechanism of both the frontline and backstage.

The circles above each represent an interface. The frontline and backstage control system can be roughly divided into four units:

- *Cashier counter***.** The cashier counter is in charge of the switch information for all interactive tables. When the customers arrive the counter will confirm the table number and switch it on. They will switch it off when the customer leaves and report back relevant information.
- *Kitchen.* The kitchen is in charge of the receipt of orders from the interactive order interface, which will be printed out directly by the printer. Once the meals have been prepared the kitchen can use the same interface to notify frontline staff to retrieve them.
- *Backstage control.* Backstage control is set up in the office of the management and is in charge of most restaurant management, including human resources, work attendance, ingredients and consumables, meal management, and customer feedback.
- *Interactive tables.* The 22 interactive tables that sense the diners' touch commands and send them to the corresponding units.

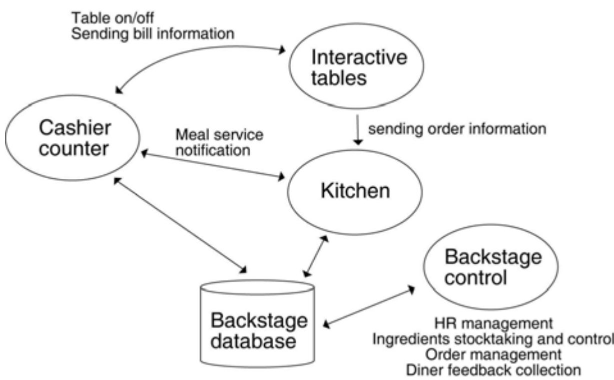


Fig. 3. Framework of the backstage control mechanism

3.6 Implementation

We installed Zytronic 22-inch ZYBRID projected capacitive touch-panels beneath the surfaces of the dining tables to sense the touch signals from diners through the table and transform them into corresponding X,Y coordinates for computer mouse cursor, allowing the system to understand where the diner touched the tabletop.

The visual interface was implemented using Adobe Flash AS3.0 combined with mySQL+PHP and connected to a database. When diners entered a command, such as placing an order, through tabletop, the system immediately recorded it in the database, connected to the printer in the kitchen, and printed the order out for the cooks, who then prepared the food accordingly.

We employed external file reading to allow restaurant owners to change tablecloth templates and the menu in the future.

4 Results

We installed 22 interactive menu tables with projected tabletops in a restaurant. After iterative closed beta tests, the restaurant was opened to public in October 2009 at 8F, No.1, Section 4, ZhongXiao East Road, Da-An District, Taipei, Taiwan, one of the most prosperous fashion and business districts in downtown Taipei because the restaurant is targeted at young people fond of fashionable items. Our unique interface design with its two-orientation display and shared circular menu worked quite well. A typical service flow at mojo iCuisine and findings derived from our empirical studies are as follow.

4.1 A Typical Service Flow

Based on observation, a typical workflow of our system's service is summarized as follows:

1. Diners enter mojo iCuisine.
2. A server greets the diners, confirms the number of people, and then seats them after switching the table to interactive mode.
3. After diners are seated, the server introduces and explains how to use the interactive tabletops.
4. Diners browse the menu using the interactive tabletop and order.
5. The orders are transmitted to the kitchen. The cooks start to process the order.
6. While their food is being prepared, diners are able to explore other features of the interactive tabletop, such as games or the changeable tablecloth.
7. The ordered dishes and/or drinks are served to the diners.
8. The diners enjoy their dishes and/or drinks.
9. The diners ask to pay the bill using the interactive tabletop. The server is signaled by a monitor at the cashier counter and goes to them.
10. The server takes the diners' money and gives change. During this time diners are encouraged to fill out the opinion form via the interactive tabletop.

4.2 Preliminary Findings

Based on our observations, the results were beyond our expectations. Mojo iCuisine seems to have advantageous capabilities not found in ordinary restaurant settings:

- *Commercial effects.* From its opening in October 2009 until June 2010, hundreds of blog posts and discussions on Facebook as well as a great deal of media exposure in Taiwan, was generated without formal marketing efforts. In addition, 9 television interviews (such as [17, 18, 19]) were conducted, 3 magazine articles were written, and the space was rented for commercial events (e.g. a company product launch) 7 times. The owner indicated that the cost was recovered within three months of the restaurant's opening, and because of its success the owner hopes to develop more, and continuously update the, digital content to boost the customer appeal and customer return rate.

- *An experimental commercial application platform.* In March 2010 when we revisited mojo iCuisine, we have found that the tablecloth had been changed to beverage advertisements, which means the restaurant made use of the tablecloth feature for commercial purposes. In addition, the restaurant provided a customized tablecloth service to diners, whereby a picture file could be uploaded to the restaurant Server beforehand and used while dining time. The restaurant owner told us that a diner once used this feature to propose marriage by replacing the tabletop graphics with a self-made proposal picture at a certain time to surprise his partner. Diners showed interest in more entertaining games, the addition of a cab-calling service, and a credit card payment feature. Another diner suggested that the cooking in the kitchen be shown live on the tabletop would be a good idea.
- *An attraction other than food.* Through on-site interviews and blog posts we discovered that the majority of the diners came to the restaurant to try out its interactive features rather than its food. Most blog articles mentioned the novelty and fun of interactive ordering, despite a few complaints that the food was not as good as expected. People may have liked mojo iCuisine more because the novel touch-sensor tables were entertaining and the atmosphere cool, and not considered the food to be their main concern. People were willing to try something new because of a potentially interesting dining experience.
- *Dual service action.* From our observations we found that after the server seated the customers operating instructions for the interactive menu were given. Once they were told the tabletop was touch-sensitive, people tended to touch here and there to explore interface features spontaneously. Except for some technical problems that occurred during early operations after restaurant's grand opening, diners were self-served successfully. The combined actions of the server and the interactive interface caused the element of entertainment to overshadow usability. Overall, customers' evaluations indicated that the dining experience was refreshing and interesting, and that they would recommend it to others.
- *Sensual spatial experience.* In the interior setting, lights were dimmed to allow better projection, thus creating a captivating indoor digital landscape. Each table glowed with a vivid digital tablecloth. Many diners took photos of the interactive table and restaurant interior and shared them on the Internet. According to some blog posts found by searching the keyword "mojo iCuisine", we found that some diners used words such as "fantastic", "magical", "fancy", "romantic", "surreal", and "cool" to describe the ambience. Some were even amazed that the "lighting table" was touch-sensitive, which we see as evidence that the ambient design successfully blended into architectural space.

5 Conclusion

This project included the design, implementation, and customer response to a digital interactive menu in a real restaurant context. This paper provides an example of an interactive system in a move towards an interactive restaurant, as well as an interactive design overhaul of the service model, which is beneficial to further studies.

The use of interactive dining tables in a restaurant was merely the first step towards an interactive restaurant, but we were already able to see the possibilities for how interactive technology could shape life styles and its influence on the integration of technology with or a service or specific activity. The database also allowed the owner to track diners' orders and keep an operating log to better understand better the restaurant's customers, which facilitates the continuous improvement of service and design.

In addition, customer creativity, such as using the tabletop to propose marriage, also helps interaction designers to discover more design themes and possibilities. In other words, the concept of combining the database with service design shows how interactive design can help the industry to continuously evolve.

This project is limited in many aspects. In terms of interface design, it is confined to a two-seat table setting. Besides, only one type of restaurant and one type of service procedure were used in a diverse dining culture with many service models. Also, the diners were mostly younger middle class consumers who were open to new things.

We hope to carry out more in-depth research regarding dining experience design and usability in the future using mojo iCuisine as our test-bed and uncover further insights.

Acknowledgments. The authors would like to thank Chu-Hsiang Chen for the tabletop graphic design, and Moxie Design Co. for interior design and hardware installation.

References

1. Weiser, M., Brown, J.S.: Designing Calm Technology. PowerGrid Journal 1(1) (1996)
2. Rekimoto, J.: SmartSkin: An Infrastructure for Freehand Manipulation on Interactive Surfaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Changing our World, Changing Ourselves (CHI 2002), pp. 113–120. ACM, New York (2002)
3. Pinhanez, C.: The everywhere displays projector: A device to create ubiquitous graphical interfaces. In: Abowd, G.D., Brumitt, B., Shafer, S. (eds.) UbiComp 2001. LNCS, vol. 2201/2001, pp. 315–331. Springer, Heidelberg (2001)
4. Paradiso, J.A., Hsiao, K., Strickon, J., Lifton, J., Adler, A.: Sensor Systems for Interactive Surfaces. IBM Systems Journal 39(3.4), 892–914 (2000)
5. Hsieh, C.-K., Liu, I.-L., Lin, Q.-P., Chan, L.-W., Hsiao, C.-H., Hung, Y.-P.: Treasure transformers: Novel interpretative installations for the national palace museum. In: Huang, F., Wang, R.-C. (eds.) ArtsIT 2009. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol. 30, Part 2, pp. 112–119. Springer, Heidelberg (2010)
6. Juan, P.H., Benjamin, B.B., Allison, D., Gustav, T.: KidPad: Collaborative Storytelling for Children. In: CHI 2002 Extended Abstracts on Human factors in Computing Systems (CHI 2002), pp. 500–501. ACM, New York (2002)
7. Michael, A.E., Jochen, R.: Collaborative Learning with Interactive Surfaces: An Interdisciplinary Agenda. In: Gomez, K., Lyons, L., Radinsky, J. (eds.) Proceedings of the 9th International Conference of the Learning Sciences, ICLS 2010, vol. 2, pp. 505–506 (2010)
8. Kurland, L.G., Gilbert, E.: Integrated Interactive Restaurant Communication System for Food and Entertainment Processing. United States Patent 4553222 (1985)

9. Chang, C.-w.: MenuVista, http://www.seeitny.com/web/img/MenuVista_01.mov
10. Inamo restaurant, http://www.inamo-restaurant.com
11. uWink, http://www.uwink.com
12. Conceptic, eMenu, http://www.conceptic.com.au/ForRestaurants.htm
13. Craig, W., Hiroshi, I., Andrew, D., Matt, G., Scott, B., Brygg, U., Paul, Y.: Ambient Displays: Turning Architectural Space into an Interface between People and Digital Information. In: Yuan, F., Konomi, S., Burkhardt, H.-J. (eds.) CoBuild 1998. LNCS, vol. 1370, pp. 22–32. Springer, Heidelberg (1998)
14. Yuan, F., Geißler, J., Holmer, T.: Roomware for Cooperative Buildings: Integrated Design of Architectural Spaces and Information Spaces. In: Yuan, F., Konomi, S., Burkhardt, H.-J. (eds.) CoBuild 1998. LNCS, vol. 1370, pp. 4–21. Springer, Heidelberg (1998)
15. Taysheng, J.: Advanced Ubiquitous Media for Interactive Space: A Framework. In: Computer Aided Architectural Design Futures 2005, vol. 6, pp. 341–350 (2005)
16. Luigina, C., Liam, B.: Space, Place and the Design of Technologically-Enhanced Physical Environments. In: Spaces, Spatiality and Technology. The Kluwer International Series on Computer Supported Cooperative Work, vol. 5, pp. 217–232 (2005)
17. Mojo iCuisine on EraNews, http://www.youtube.com/watch?v=lGy0yY7srms&fmt=18
18. Mojo iCuisine on CTV, http://www.youtube.com/watch?v=-DutolIRNk0
19. Mojo iCuisine on FTV, http://www.youtube.com/watch?v=iuxZfpaTZw0

Usability of Nomadic User Interfaces

Walter Dees

Philips Research Laboratories, High Tech Campus 34, Eindhoven, The Netherlands
walter.dees@philips.com

Abstract. During the last decade, a number of research activities have been performed to enable user interfaces and the underlying user activities to be migrated from one device to another. We call this "Nomadic User Interfaces". The primary goal of these research activities has been to develop the technologies to enable this. However, not much is known about the usability aspects of Nomadic User Interfaces. In this paper we present the results of three different user tests that we conducted to investigate the usefulness and the usability issues of several prototype Nomadic User Interface systems that we developed.

Keywords: Nomadic User Interfaces, UI Adaptation, Migratory Interfaces, Pervasive Applications, Ubiquitous Access, Device Independence, Multi-device UI Authoring, UI Consistency, Session/Activity Migration, Session Mobility, Multimodal Interfaces, Remote User Interfaces, User Study, Usability Guidelines.

1 Introduction

Interconnectivity of devices using technologies such as Bluetooth, DLNA, and broadband allows new usage scenarios. One of the opportunities enabled by the enhanced interconnectivity offered by these technologies is to give people more flexibility in when and where to access their favorite content and applications. The next step in enhancing these types of scenarios is through a concept that we call "Nomadic User Interfaces" (also known as "Migratory User Interfaces" [1]). The main idea behind this concept is that interactive user interfaces offered by an application in the network can be freely migrated from any stationary and portable device inside the network to other devices in the network. It allows users to start an activity on one device and continue this activity on another device. This is illustrated in Figure 1.

Examples of Nomadic User Interface scenarios may include a user creating a playlist on a tablet device, and then start playing the contents of this playlist on the living room TV and/or stereo system. When the user goes to bed, he can continue to finish the playlist in the bedroom, skip a few items and stop the playlist through a user interface on the bedroom TV. Another example could be a user typing a message on a mobile phone and continue typing the message on a big screen TV with a wireless keyboard.

The topics underlying Nomadic User Interfaces, such as UI adaptation techniques, model based design of user interfaces, (semi-)automatic generation of multi-device user interfaces, session migration techniques and remote user interface technologies

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 195–204, 2011.

have been a topic of research for several years. Examples of research results in this area include: TeresaXML/Migration project [2], UsiXML/MigriXML [3], UI Plasticity [4], Supple [5], and Multi-level Stylesheets [6][7].

Fig. 1. Nomadic User Interfaces

The primary goal of most of these research activities has been to get the technologies in place to enable concepts such as Nomadic User Interfaces. Not much research has been done to measure the usability criteria of Nomadic User Interfaces and to identify the usability issues of such a system in practice. Therefore, we have implemented several prototypes, and conducted three different user tests to investigate this.

2 Study 1: Exploratory Study

2.1 Introduction

The first user study was a qualitative study on the usefulness and usability issues of Nomadic User Interfaces, with a focus on in-home usage. The prototype created for this study allows migration of the GUI of a media browser application to an HD and SD resolution television, mobile phone, PDA, tablet PC and an iPronto remote control. The prototype supported both push- and pull-migration. This means that a user can instruct the system to move (i.e. “push”) the current activity to another device, or at any moment use another device to discover a list of currently active and suspended activities, and continue (i.e. “pull”) the activity to the device that the user is currently using.

2.2 Setup of the Experiment

The experiment was conducted at the Philips ExperienceLab [8], and involved eight sessions of approximately two-hours in length. The participants of the experiment were 8 couples (mostly married or couples living together). Their familiarity to networking concepts ranged from high understanding of the underlying technical mechanisms to no prior experience of using networked systems, while there was a bias towards educated and financially fortunate people. However, this bias was not considered as misleading since such users seem to be the initial target group of the Nomadic User Interface concept.

After a short introduction of the experiment and the system, the users were given complete freedom to transfer their activities from one device to another, to different devices located throughout the home, i.e. a living room TV, a bedroom TV, a touch screen display in the kitchen, a tablet, a mobile phone and a PDA. During the exploration phase the experimenter was present, observing the users at a distance, only interrupting the participants when absolutely needed. Subsequently, certain role playing scenarios were employed, such as watching pictures that were stored on a mobile phone together on a TV, creating a playlist, and moving a session from the TV downstairs to a TV upstairs. The participants were asked to “think aloud”, in order to observe their thoughts and behaviors. The entire two-hour session was recorded using the camera infrastructure of the Philips ExperienceLab, for further analysis.

Fig. 2. Views from the experiment

2.3 Analysis of the Results

As mentioned, the prototype supported both push- and pull-migration. The pull-migration was considered to be the most favorite of the two. The reasons that were given for this are that the push-migration paradigm entails premature commitment while in-home interactions are rarely planned. It also is unclear at the moment they push the activity to another device in another room, whether or not that device is actually available (e.g. switched on, not occupied by someone else). Furthermore, push-migration does not deal with the issue of the “dead time” during the relocation. Users mentioned for example that they may want to transfer their current activity to the bedroom TV, but first want to go and brush their teeth. People want to be sure that the status of their session will be kept intact. Although the prototype supported automatic suspension of the activities, this does not give sufficient feel of control. It was felt that it was removing control from them, and that they rather have an explicit action to suspend the UI (before migration) than automatic suspension by the system. The “suspend” or “pause” metaphor is very familiar to the users and therefore should be considered for this type of scenarios. Also, during the study people expressed a need for having the ability to mirror and split the UI across multiple devices, next to pull and push migration.

One of the main issues that this study revealed is that the notion of networked activities is not readily comprehensible by all users. They were confused as to where the session was located, and felt a need to have some indication with which device they are communicating and where the data comes from. During the experiment we also noticed a hesitant attitude towards the usefulness of migrating activities that involve a high degree of interactivity. This may be due to the perception that users have about each of the individual devices, and the capabilities/limitations of interacting with some of these devices. This requires further study.

Based on the feedback from the experiment, we were able to classify some of the different user needs relevant to the concept of Nomadic User Interfaces. Figure 3 shows a possible classification of user needs for Nomadic User Interfaces in domestic environments. Next to the mobility of activities across devices, triggered by certain interruptions, it also shows the relevance of Nomadic User Interfaces in cases where an activity is started on one device, but a device which offers enhanced interaction is available in the vicinity, e.g. moving the activity from a mobile phone to a large screen TV.

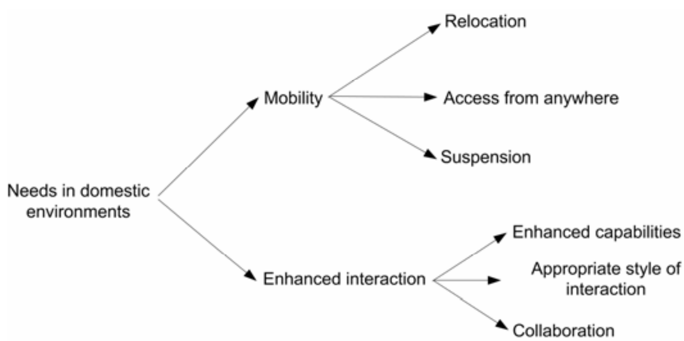


Fig. 3. Classification of user needs relevant to Nomadic User Interfaces

3 Study 2: Deeper Analysis

3.1 Introduction

The main focus of the second user study was to gain a deeper insight in the user needs and usability requirements regarding Nomadic User Interfaces, and in particular different forms of migration and session state recovery. In order to test this, a new prototype was implemented that supports the following types of migration:

– The first is *total migration*, whereby a running activity can be moved from one device to another in its entirety. This can be done using either a push mechanism or pull mechanism, as described in the previous section. Once the current activity is transferred to the new device, the device from which the activity was moved becomes passive and the new device becomes active. The user starts interacting with the new device using available interaction means of the new device.

- The second is *partial migration* and in this case a running activity is transferred partially from one device to another. This distributes or splits the user interface over two devices. For example, once the user discovers a large screen TV in the vicinity, the user moves the activity that is running on the mobile phone to the TV, whereby the control part remains on the mobile device and the presentation part is moved to the TV screen. An example of this is given in Figure 8.
- The third is *mirroring* (also known as *cloning*), whereby the UI is shown on both devices simultaneously and both UIs are kept in sync.

Furthermore, for both partial and total migration different types of state recovery granularity, such as session level, task level, and action level granularity (as defined in [9]) were supported by the prototype. The main target devices of this second prototype were a mobile phone, a TV and a PC. A photo manager/sharing application, with which the users could sort their photo collection, annotate pictures, create albums and slideshows, and share photos with friends, was chosen as the carrier application.

3.2 Setup of the Experiment

In total 5 users, all well-educated people, participated in sessions of 1.5 hours each. The study was a controlled user study conducted at the Philips ExperienceLab whereby users had to follow different tasks in a task book. These tasks were steps inside a larger scenario that included various forms of migration due to different kinds of interruptions (e.g. receiving a message from a friend whilst organizing pictures). Some tasks had to be repeated several times in order to get some insight into the preferred state recovery granularity.

The user had to fill in a questionnaire each time a task was completed. There was also a pre- and post-test questionnaire which included more general questions about session migration and its usefulness.

3.3 Analysis of the Results

The following figure shows the result of some of the questions in the questionnaires that were filled out.

These results show that the participants appreciated the session migration feature, and found it useful to have such feature. We also asked the participants some questions about which activities they currently perform on more than one device and on more than one location. In total, the participants suggested 11 different activities with which they currently have or would like to have a continuous involvement with across devices and locations, i.e. activities for which a session migration feature could be very useful. Chatting and gaming were amongst the most frequently suggested.

Regarding the question about whether people prefer total migration, partial migration or mirroring when moving an activity from mobile phone to TV, people ranked them as follows: partial migration was clearly preferred by 3 users, cloning was preferred by 1 user, and push migration by 1 user. Unfortunately, the number of participants for this experiment was very low, which makes these results inconclusive.

The test revealed that we should always aim to recover as much of the UI state information as possible to ease the interaction continuity. The effect does however become less the longer it takes between a session save and a session restore.

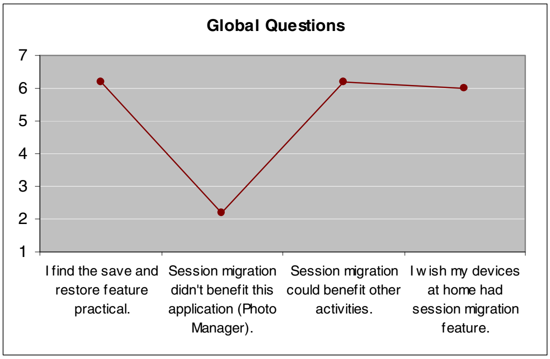


Fig. 4. Results of global questions questionnaire of second user test using 7-point Likert scale

4 Study 3: Consistency of Multi-device Interfaces

4.1 Introduction

The main focus of the third study was to investigate the consistency requirements of multi-device user interfaces in a controlled manner. Research questions included: Which consistency parameters are most relevant when moving a session from one device to another for the user to not get lost. Do people prefer intra-device or intra-application consistency? Can we define some consistency and multi-device UI authoring guidelines for Nomadic User Interface systems?

For investigating these questions, the second prototype was altered a bit and a number of variations in the interface and interaction (both consistent and inconsistent) were introduced into the system. There were some intra-device inconsistencies and inter-device inconsistencies. Globally, six different types of interface and interaction variations were implemented for two different devices: mobile phone and TV. These interface/interaction variations included differences in look&feel parameters, layout,

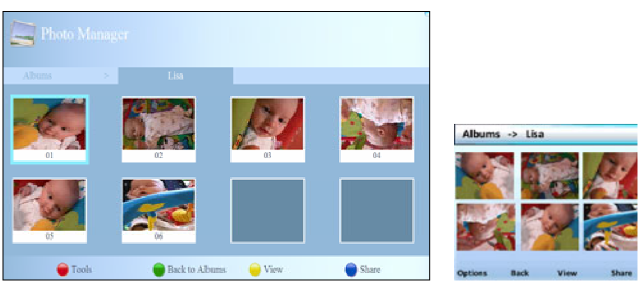


Fig. 5. Example of two corresponding user interfaces

widget type, and touch/key-based interaction. Figure 5 shows an example of two corresponding user interfaces with a consistent colour scheme, but with different interaction style.

4.2 Setup of the Experiment

In total, 24 users participated in this user study. A proper participant screening was done before confirming participants for the study. We tried to have both expert users (who are familiar with some of the technologies and devices used) and novice users (who are not familiar with the technologies and devices used). The research model in Figure 6 shows which attributes were controlled and which were manipulated during the experiment.

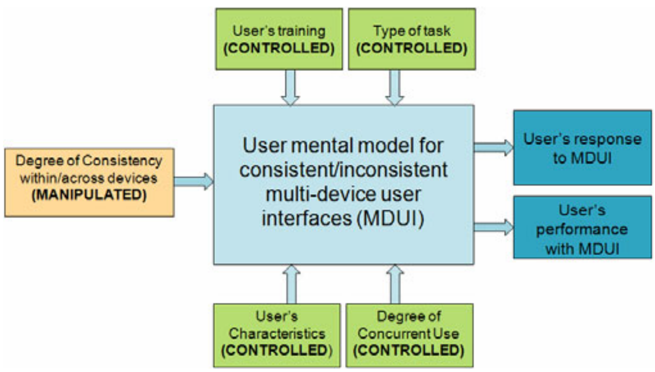


Fig. 6. Research model used for experiment

4.3 Analysis of the Results

The results of the usability study gave a good insight into the user's mental model that they constructed when they moved between two consistent user interfaces and inconsistent user interfaces. It also revealed which consistency parameters are more important than others. The results reveal that one of the most troublesome inconsistencies is interaction inconsistency (e.g. from remote control interface to touch interface). The impact of color inconsistency was also high. The use of consistent colors across interfaces running on two different devices gives a feel of connectedness and users perceive such an application running on two different devices as a single application.

The results revealed that consistency is an important attribute for enhancing the overall user experience (see Figure 7).

However, it is also learned that consistency is not the only requirement to ensure cross-device usability and in some cases it is important to sacrifice consistency for getting the benefits of local platform capabilities. For example partial migration, whereby the user interface output is rendered on the TV whilst the user input is done on the mobile phone, is very inconsistent (see Figure 7), but has low cognitive overhead, and was perceived as a very natural way of interacting. Also, during some of the tasks, users explicitly chose and preferred using touch screen based interaction instead

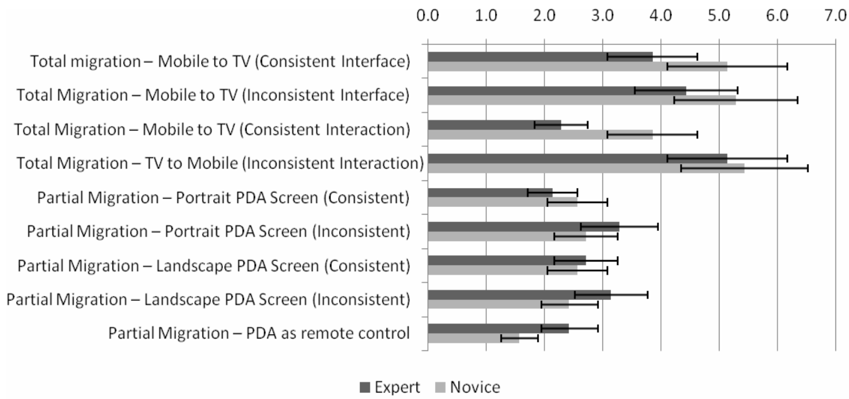


Fig. 7. Cognitive overhead of session migration (lower score means less overhead and therefore easier to use)

Fig. 8. Partial migration

of keypad based interaction although keypad-based interaction was more consistent to the remote control based interaction used for TV.

Based on the results of this usability study, several multi-device UI authoring guidelines were formulated. These guidelines can assist designers in transforming one platform specific interface into another and to generally improve the UI authoring process for multi-device user interfaces. Examples of such guidelines include:

– Wherever possible, design for cross-platform consistency instead of local (i.e. native) look and feel consistency, in particular maintain cross-device consistency of color.
– Make sure that the "point of focus" of the user is kept consistent across migrations.
– Scale items instead of repositioning or re-layouting the items (e.g. by adding whitespaces) when moving to a different screen size.
– Partial migration requires a carefully designed UI that is consistent with the interaction style and menu style of the input device, but also consistent with screen layout of the output device.
– Instead of renaming long labels of a widget, apply truncation rules (e.g. replace "accept changes" with "accept…" or "accept chan…") when moving to a smaller screen size. For normal text, it is better to define a set of text transformation rules where long text can be replaced with semantically equivalent short text.

– Whilst replacing a widget which has more than one item with another widget, do not change the order of items or item selection colour.

We also analyzed the consistency impact of various visual element manipulations. Figure 9 shows which differences between two user interfaces are considered more harmful than others, when moving from one platform to another. The changes on the top are less harmful and are thus perceived as less inconsistent. Changes at the bottom are more harmful and are perceived as inconsistent.

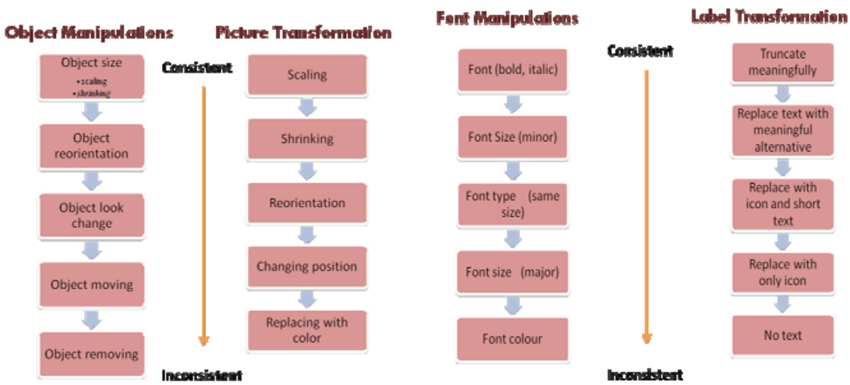


Fig. 9. The impact of different manipulations on the perceived consistency

5 Conclusions

The three studies as presented in this paper provide insight in some of the user interface needs and issues behind the concept of Nomadic User Interfaces.

The first user study showed that people like the pull-migration paradigm over the push-migration paradigm. Furthermore, the study revealed a need for users to have explicit control over suspending/pausing the user interface, rather than the system doing this automatically. The study also allowed us to get better insights in the user needs of Nomadic User Interfaces (such as mobility and enhanced user interaction offered by another device), and some of the practical problems and issues.

During the second study we gained more insights into which type of migration (i.e. total migration, partial migration or mirroring) and which type of state recovery granularity people prefer when moving an activity from one device to another. However, due to the low number of participants the results of this test were rather inconclusive, although the study did confirm our expectation that we should always aim to recover as much of the UI state information as possible to ease the interaction continuity.

The third study revealed the requirements related to the consistency of user interfaces across devices, when used in a Nomadic User Interface context. It showed that consistency is a very important factor in usability of multi-device user interfaces. However, the results also show that it is not the only factor. Based on this study, we formulated a number of design guidelines that will help to improve the UI authoring process for user interfaces that span multiple devices or that can move freely from one device to another.

Further studies are needed to get a better understanding of the impact of the capabilities/limitations of certain devices on the willingness to migrate certain activities from one device to another. Also several next steps would have to be made in improving the Nomadic User Interface framework, its user interaction concepts, and also the technologies and tools to create Nomadic User Interfaces, in order to see if Nomadic User Interfaces really has the potential to become successful.

The results of these user tests, including the problems and issues that have been identified during these studies, are therefore valuable input for further research and future deployments in this area.

Acknowledgements. This research could not have been done without the help and valuable support of P. Shrubsole and R. Haakma from Philips Research, Prof. Dr. P.M.E. de Bra, Dr. E. Karapanos, and B. Golsteijn from Eindhoven University of Technology, S. Shahid from Tilburg University, Prof. Dr. L. Hardman from CWI Amsterdam, N. Frame from the University of Bordeaux, and J. Coutaz and G. Calvary from the University of Grenoble.

References

1. Berti, S., Paternó, F., Santoro, C.: A Taxonomy for Migratory User Interfaces. In: Gilroy, S.W., Harrison, M.D. (eds.) DSV-IS 2005. LNCS, vol. 3941, pp. 149–160. Springer, Heidelberg (2006)
2. Migratory User Interfaces, `http://giove.isti.cnr.it/tools/Migration/`
3. UsiXML - USer Interface eXtensible Markup Language, `http://www.usixml.org/`
4. Coutaz, J.: User interface plasticity: model driven engineering to the limit! In: Proceedings of the 2nd ACM SIGCHI Symposium on Engineering Interactive Computing Systems, EICS 2010, pp. 1–8. ACM, New York (2010)
5. Gajos, K., Weld, D., Wobbrock, J.: Automatically generating personalized user interfaces with SUPPLE. J. Artificial Intelligence 174, 910–950 (2010)
6. Dees, W.: Handling device diversity through multi-level stylesheets. In: Proceedings of the 9th International Conference on Intelligent User Interfaces, IUI 2004, pp. 229–231. ACM, New York (2004)
7. Dees, W.: Device Independent Layout and Style Editing Using Multi-Level Style Sheets. In: Proceedings of the Sixth International Conference on Computer-Aided Design of User Interfaces V, pp. 183–190. Springer, Heidelberg (2007)
8. Philips Research - ExperienceLab, `http://www.research.philips.com/focused/experiencelab.html`
9. Balme, L., Alvaro, X., Calvary, G., Demeure, A., Sottet, J.-S.: An MDE-SOA Approach to Support Plastic User Interfaces in Ambient Spaces. In: Stephanidis, C. (ed.) UAHCI 2007 (Part II). LNCS, vol. 4555, pp. 63–72. Springer, Heidelberg (2007)

Adaptive Implicit Interaction for Healthy Nutrition and Food Intake Supervision

Felix Kamieth, Andreas Braun, and Christian Schlehuber

Fraunhofer Institute for Computer Graphics Research IGD, Fraunhoferstraße 5,
64283 Darmstadt, Germany
{firstname.lastname}@igd.fraunhofer.de

Abstract. The current work is going to provide you information about our solution in the challenge of nutrition and food intake supervision, which has been developed lately. We will give an overview of the system and the implemented mechanisms, which were needed for aiding users in supervising and improving their eating habits. We will show the features, which may be useful for persons who want to analyze their eating habits and try to improve those. Therefore our system provides a cooking advisor, which is able to recognize the available food and respecting those presents the user a list of recipes, which fit his available ingredients and also his nutritional needs. If he wishes, he has also the possibility to set other filter parameters. Additionally the cooked menus are logged by the system and may be subject to further analyses. For determining the available ingredients our system uses RFID technology and also provides the user some community-like features for submitting new receipts or new ingredients.

Keywords: Food, Nutrition, Supervision, RFID, Cooking Advisor, Food Intake, Healthy Nutrition.

1 Introduction

Adequate healthful nutrition is growing as a widespread health-concern in the western world. The growth of prepared meals and fast-food consumption has reduced the healthfulness of the common western diet. To help with the cooking process only few solutions exist. They mainly focus on the cooking itself and due to that try to help the user during the process by for instance reading out the selected recipe and giving advice on how to perform the needed cooking actions accurately. However, in the process of finding an accurate recipe the user is left on his own.

Therefore this paper describes a solution, which helps the user in the process of finding an adequate recipe for his meal, without having to read dozens of recipes, were he realizes, that always some ingredients are missing in his stock.

Section 2 will give a short introduction to the area of cooking advisor and food intake supervision systems and describe related work. Section 3 describes the general design of the system itself and details the three central tasks. Section 4 focuses on the interaction and the user interface of the system discusses and the considerations which led to the current design. Section 5 analyses the conclusions drawn from the research and discussed future areas of work.

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 205–212, 2011.

2 Related Work

In the area of cooking aid and eating supervision several approaches has been followed. However, most of these focus on the process of cooking itself. In [2] the authors present a system, which offers information about the process of cooking a desired meal and guides the user step-by-step through this process, while it analyzes the user's actions and warns him of cooking errors.

Other work [1] focuses completely on recipe finding for diverse conditions in which a human being can be. This means that the system analyzes the user's preferences and gives every meal in its database a score according to the complexity of the cooking process. In a final step the user is presented a list of recipes, ordered by ease of cooking, which fit to his actual condition.

In contrast to the former mentioned works the system described in this paper does neither provide any corrections regarding the user's actions during the cooking process, nor does it automatically analyze the user's condition and takes it into consideration. But it has a great benefit in comparison to the both existing solutions. While the user in the both presented solutions gets a list of recipes and then has to look if he has the needed ingredients, our system takes the available ingredients into consideration and only presents the user the recipes, which he really is able to cook.

The "intelligent" refrigerator, which is presented in [3] is close to our work. The authors make use of RFID technology for determining the content of the fridge and based on this content list the fridge then gives the user suggestions for possible meals. The possible recipes are gathered by the fridge by communicating with other "intelligent" fridges in the neighborhood.

This solution has some parallels to our work, but in contrast to this work we do not limit the scope of gathering ingredients to the fridge, but on all food, which is used and we also provide food intake supervision and take user set restrictions into consideration when resolving possible recipes.

3 System Architecture

The presented system can be divided into two stand-alone parts, the client, which is present at the computer of the user and the server application, which is fed by the clients with new recipes and hands out updates to the clients. The main tasks to be completed by the presented system are:

- Gathering the available ingredients and finding appropriate recipes which fit the users nutrition criteria
- Collecting new recipes and ingredients, which can be contributed by the user
- Logging and analyzing the eating habits of the user.

For completing these tasks our client-system has to consist of a database subsystem with an input device, like a barcode scanner or a RFID reader, for recognizing the ingredients. The interaction between those elements is illustrated in figure 1.

The next subchapter outlines the representation of food as a data object and provides a more detailed description of each of the previously presented tasks of the system.

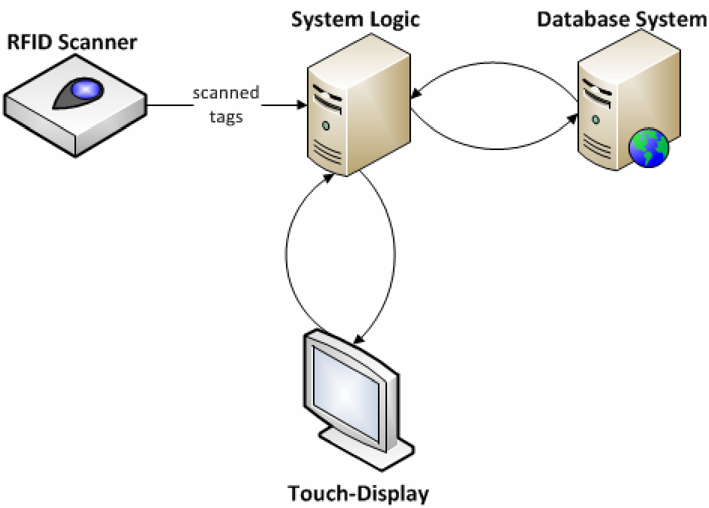


Fig. 1. Interaction between HW and SW components

3.1 Representing Ingredients and Meals

Before being able to start with searching for recipes or even supervising the food intake of a user the representation of food and recipe has to be defined.

Every kind and brand of a food has to be taken into account, while it is also possible to exchange different kinds of an ingredient with one of another brand in a recipe. For respecting all of these constraints each food element is applied with a field for the brand and also the type.

In contrast to the challenge of representing a single food ingredient a recipe is much easier to represent in a data structure. It consists of many ingredients with a specific amount of each of those ingredients and also can be an ingredient itself (for instance you may want to only eat a salad or eat it as a starter for a meal).

The database model can be seen in figure 2.

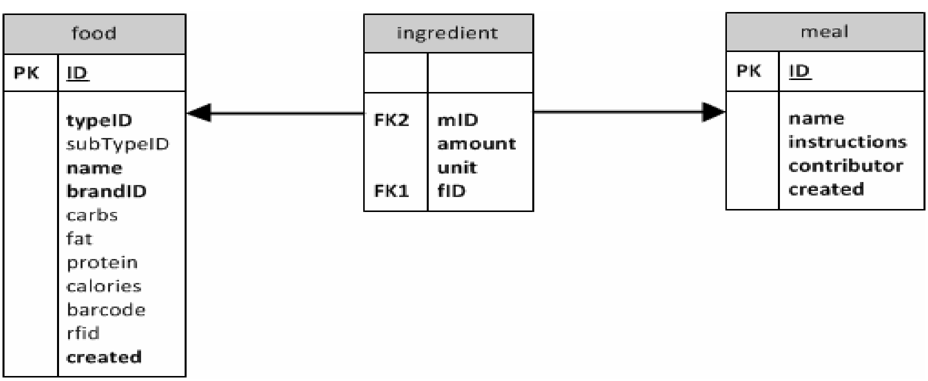


Fig. 2. Extract from the database representation of food and meal

3.2 Gathering of Available Ingredients and Recipe Finding

This subchapter describes the ingredient gathering process as well as the method of finding recipes used in this system. It also outlines the means for the collection of new recipes and ingredient descriptions.

Finding available ingredients. Entering available ingredients into the system the user has three possibilities:

1. Scanning the barcode
2. Entering the barcode by hand
3. Scanning the RFID tag.

The first two variants are based on the barcode, which is currently present on most commercial products found in stores. The second variant is included to expand the use of the system to users who do not own a barcode or a RFID scanner.

In contrast to the first two variants the scanning via RFID is more comfortable, because the user does not have to scan each new bought ingredient on its own, but the whole basket can be placed near the reader and the tags will be read automatically.

At the end of the input process the system will add the new bought items to the existing stock of known ingredients.

Resolving all possible recipes. After the system has a list of all available ingredients present the user in a next step has the option to set some additional search parameters, for instance he only searches for vegetarian food. If everything is as he desires he has the possibility to start a recipe search run.

The search process works in two phases. In the first phase the system searches for direct matches, which mean that the ingredients in the recipe have to match exactly the available ingredients (brand, type, etc.). After this the result is shown to the user and if he is already satisfied the search is finished at this point.

If no result has been resolved or if the user is not satisfied a second search phase is started, in which the ingredients are abstracted to their types, so that the search matches recipe ingredient types to available ingredient types. The process can be seen in the listing below.

```
...
// $deepSearch = user set search flag
// $allRecipes = a list containing all known recipes
// $avIngredients = a list with all available ingreds

$filtRecipes = filter($allRecipes, $userFilterParams);
$result = directSearch($filtRecipes, $avIngredients);
If($result is empty || $deeperSearch){
$result = typeSearch($filtRecipes, $avIngredients);
}
...
```

3.3 Collecting New Recipes and Ingredients

For establishing a broad database and keeping up to date with new recipes or ingredients a community feature is included in the system, which enables each user of the client to submit his self-entered recipes or ingredients to an online server, so that all other users can benefit from the user's work.

For doing so he is given the option of entering all available nutritional data along with the name of the product and some other identification relevant data on an input window. If he entered an ingredient he besides this information also has to scan the RFID tag applied to it or the barcode.

The new items are at first stored in the local database and if the user enabled the upload of his created items these are committed to the server during the next synchronization process.

Synchronization process. The synchronization process is started at a user-defined time or by manually starting the update of the client. During this process the recipes and ingredients the user wants to contribute are uploaded to the server. After successful upload, these are marked as "submitted", so that future synchronizations will only upload new entries.

Besides this, the client submits a timestamp to the server, which tells it, when the last update of the client's recipe and ingredient database has taken place. The server then gathers all recipes and ingredients which have been included into the online database since this date and sends these entries to the client.

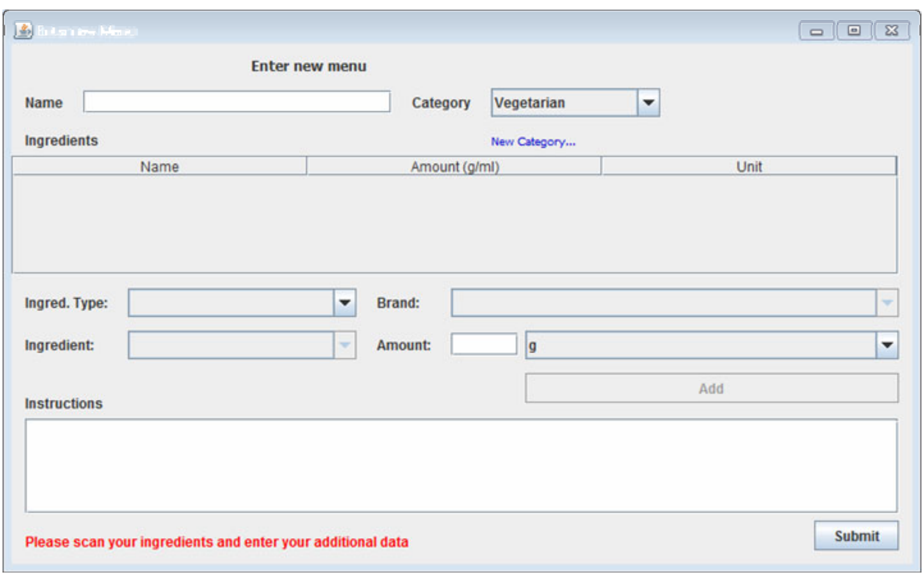


Fig. 3. The input dialog for creating a new menu

3.4 Logging and Analyzing the Eating Habits

After searching a recipe the user will normally start the cooking process. He then has the possibility to report the system that he has cooked or is cooking the meal now. The system then will calculate the calories and other nutritional data from the chosen menu and log them into the internal database together with the time and the identifier of the meal.

Due to this logging the user will be able to reconstruct his meals over a desired time. Besides this fact he has the ability to read the nutritional data from the logged list and find eventual problems in his eating habits, like too many calories over a long time, and gets the ability to correct this.

The logged list can also be exported in a comma-separated file, so that it may be imported into a statistical application for further analysis of eating habits.

4 Interaction Design

For the GUI and the interaction design of the system several things had to be considered. The use of the system should be easily learned and intuitive - it also had to be considered, that the system may be used by disabled or elderly people.

4.1 UI Structure

The user interface of the software is split into two groups. First is the standard presentation, which is used to search for an appropriate recipe. These interaction windows support touch-display use. Second is the group of maintenance related windows, which is needed for recording new ingredients or new recipes by a contributing user. The difference between these types is that the presentation windows are designed to fit the requirements of a touch display, which means bigger clickable buttons and larger interaction elements. The maintenance items in contrast are designed as input forms to allow the user to present different data to the system, when contributing a new recipe or ingredient.

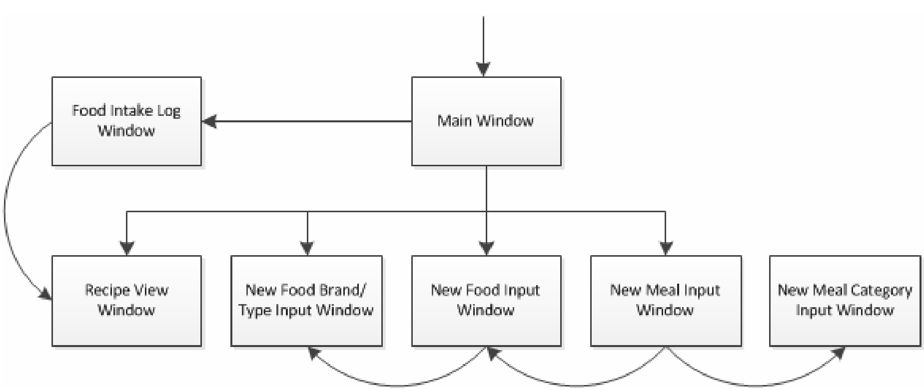


Fig. 4. Control flow between the UI windows

The control flow between the interaction windows has been designed such that the user always has a direct way to access a corresponding interaction window needed. This structure can be seen in figure 4.

4.2 UI Design

The design of the user interface has been done respecting several rules of user interface design, which apply the same for standard and modern in- and output devices.

The most important design principles were the following:

- **Visibility principle:** Due to this principle all task-relevant controls should be visible to the user from the beginning on. This becomes even more important, if a touch-display is used for interaction, because there the usage of several menus for finding an action may distract the user.
- **Structure principle:** This principle aims at building user interfaces as structured blocks, in which all interaction elements aim at one task. This makes the interface easy to understand and an intuitive usage is possible.

5 Conclusion and Future Work

In the previous sections we have shown the design of a food intake and nutrition supervision system. It has been shown how the gathering of available ingredients can be improved due to the usage of RFID or barcode scanner based systems and how these techniques can be used to aid the user in the process of finding a recipe for cooking a meal.

Currently the database of recipes and ingredients depends on the contributions of users of the system. Besides this it could be considered to develop mechanisms, which make it possible to synchronize the system with some of the bigger cooking websites currently available to get new recipes.

In addition it may be considerable to expand the features of the eating habit analyzer considering the change of nutritional data of the ingredients during the cooking process and providing the user with a better approximation of his current intake of calories, etc. Besides this the system may be able to determine the healthfulness of the eating habits of the user and give advice to improve those.

In the here provided work the user has the ability to manually filter recipes for his needs (for instance because he is a vegetarian). Future work could be made in the development of a new subsystem, which works as a real nutritional advisor. The system could be given a nutritional target by the user (i.e. losing weight). Then the advisor subsystem would analyze the meals of the user over a period of time and then provide him a diet plan for reaching his goal. The diet plan then would influence the recipe search, so that the resolved meals fit into the plan.

References

1. Yajima, A., Kobayashi, I.: Easy Cooking Recipe Recommendation Considering User's Conditions. In: International Conference on Web Intelligence and Intelligent Agent Technology. IEEE, Los Alamitos (2009)

2. Fadil, Y., Mega, S., Horie, A., Uehara, K.: Mixed Reality Cooking Support System using Content-Free Recipe Selection. In: Proceedings of the Eight IEEE International Symposium on Multimedia. IEEE, Los Alamitos (2006)
3. Gu, H., Wang, D.: A Content-aware Fridge Based on RFID in Smart Home for Home-Healthcare. In: 6th International Conference on Autonomic Computing and Communications. IEEE, Los Alamitos (2009)

Recall and Communication Support System for Reminiscences Triggered by Humming

Yusuke Kita and Yoshio Nakatani

Graduate School of Science and Engineering,
Ritsumeikan University College of Information Science and Engineering,
Ritsumeikan University 1-1, Noji-Higashi 1, Kusatsu, Shiga 525-0058, Japan
cc006061@ed.ritsumei.ac.jp, nakatani@is.ritsumei.ac.jp

Abstract. This paper proposes an effective reminder management system involving fond memory communication. Our framework helps users to manage and recollect fond memories they have. The system pays attention to the fact that unconscious humming is deeply related to the recollection of fond memories, and proposes a fond memory recollection system utilizing it. The system identifies a hummed song, and if a fond memory is related to that song and has been registered in the data base, metadata that accompanies the music gets displayed. Moreover, the system analyzes the music that the user used to trigger a fond memory recollection, resulting in recommended music that the user would probably like, along with other people's fond memories, being displayed.

Keywords: fond memories, recollection, communication, humming.

1 Introduction

People accumulate fond memories every day. Fond memories are part of what makes a person individual, while also being a form of communication by sharing similar memories and trying mutual understanding. The importance of fond memories was appealed to by Nojima, H. who advocated fond memory engineering that arranges and manages them [1]. We examined methods of managing, storing, and using fond memories, and then created a technological support framework for doing so.

This research focuses on the relationship between humming, which is an unconscious activity, and memory recollection. The system involves a memory recollection system that utilizes humming as a trigger.

Most past research has mainly used photographs, and rarely used music, in triggering people to recollect a scene from a certain time in fond memory engineering. We tend to recall past memories for no special reason when listening to music because there is close relationship between music and fond memories. Music is a strong trigger of prompting memory recollection, but an effective method of music-based fond memory communication support framework has yet to have been established.

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 213–219, 2011.

2 Background

Fond memories involve three functions: self function, society function, and cognitive science instruction function [2].

The self-function provides a person with the base for supporting his/her own consistency and maintaining a preferable self-image. It is very useful in identifying growth by comparing the past self with the present.

The social function is useful in forming and maintaining interpersonal relationships. Moreover, inserting one's own experience into a conversation can improve the credibility of and enrich communication.

The instruction function is useful in directing various judgments and actions. Recollection of similar past experiences related to a present problem can be then utilized in problem solving and planning, which has been technologically modeled as case-based reasoning in Artificial Intelligence. Furthermore, past experiences motivate people and contribute to decision of their attitude.

These three functions are very important to support fond memory recollection, and thus which function should be constantly supported by the system is necessary to be considered.

3 Related Research

Some systems that effectively use the self and instruction functions have already been proposed for use in fond memory engineering. Photographs have been mainly used up to now to trigger fond memory recollection because of their strong evocativeness [3]. However, photographs are not suitable for supporting the social function because they are generally records of personal experiences. Nakatani, Y. pointed out importance of fond memory recollection using music [4]. Popular songs are shared among people and can be representative of a certain time. They are very strong triggers that encourage fond memory recollection and confirm companionship of the same generation sharing the similar experiences. Fond memories are recorded and recollected through their relationship to both past social and personal circumstances, and a song that is unconsciously expressed by humming it can include not only personal information but also information on the background time of that fond memory, thus making it an effective trigger in fond memories recollection. Music is especially better at evocating emotion and a sense of nostalgia than a photograph, and it can encourage a sharing of the consciousness of people of the same age.

4 Overview of the System

4.1 Automatic Search of What Is Hummed

We often hum in our daily lives, with doing so often increasing the possibility of unconsciously recollecting a fond memory. If information on the song hummed were to be made available at this time, it would consciously encourage fond memory recollection. Providing tags with the music would enhance effective memory recollection via being visual triggers.

The system uses an iPhone application "midomi" made by Melodis Company to identify the name of a hummed song [5]. The name of the hummed song and its singer are provided by this application. Then the system provides metadata related to the song (see 4.3). Some of the information gained in this process are used as triggers, and thus expected to encourage the recollection of a related fond memory into a chain reaction.

4.2 Automatic Detection of Rhythm

We tend to beat out rhythms during daily activities. At that time, humming the same rhythm song can often occur. Many experiments suggest that music enhances the storage and recall of memory. When a song and an event occur at the same time, the event is stored with a link to the music. Next time when the song is played, the memory is easily recalled, thus evoking a sense of nostalgia [6]. It can therefore be used to encourage fond memory recollection by the rhythm of a typical activity being automatically detected and then the same rhythm song provided.

This system uses an acceleration sensor to automatically detect a personal rhythm. When the system detects a rhythm of a person, then it provides a user with music of a similar rhythm to the user's rhythm.

4.3 Metadata

This paper treats music as a trigger for fond memories recollection. Although the most important information included in a fond memory is an episode itself, the metadata of the episode, or when and where that episode occurred, is also important and useful to effectively manage and recollect fond memories. In this system, the metadata include classification of music, and "folksonomy" is used to classify music. Folksonomy involves classifying data without hierarchy, using tags that the user can freely define [7]. Music can be classified from various kinds of personal viewpoints specific to an individual user.

Fond memories can also be tagged by the folksonomy in the same way as music. Because more than one tag can be attached to a fond memory and a song, the fond memory and song can be classified and associated flexibly from several viewpoints. A list of all the tags is displayed on a computer screen as a form of "Tag Cloud." Tag clouds are a user interface element commonly associated with folksonomy tags. Here, more frequently used tags are depicted in a larger font (Figure 1). Largely depicted tags often remind people of related memories, so a tag could is an effective trigger, as well as a platform of classification.

Fig. 1. Image of tag

4.4 Music Recommendations via Collaborative Filtering

If two persons prefer to the same song, their preference of music may be similar, and vice versa. This suggests that memory recollection can be prompted by a song which is preferred by the other user who has similar preferences of music as the user. Based on this idea, the system recommends songs by using "collaborative filtering." Collaborative filtering is a method of recommending information that a user would probably like from a large amount of information [8].

The system makes recommendations based on the assumption as follows: if the user and another user, A, use the same music as a trigger, other music preferred by the user A could very probably be used as a trigger for the user. When users register fond memories, they are requested to register an associated trigger song. The system manages a list of all songs which were used as triggers. When a new song is used, it is added to the bottom of the list. Based on this list, the system prepares a used song list for each user. Each value of the list elements is set at 1 if the corresponding song has been used as a trigger by the user, and else 0. The system calculates the degree of similarity among used song lists, and recommends a song among the used song list of the other user with the highest degree of similarity to the users. The system uses the Pearson product-moment correlation coefficient that was used by Ringo to calculate the degree of similarity [9].

5 System Architecture

The system supports fond memory recollection, management and communication via three functions described above. PHP and Mysql were used in creating the system because PHP is very suitable for developing web applications and MYSQL is compliant with PHP. The current system is implemented on the mobile PC and will be implemented for use on iPhones in the next step because it is assumed it will be used in daily life.

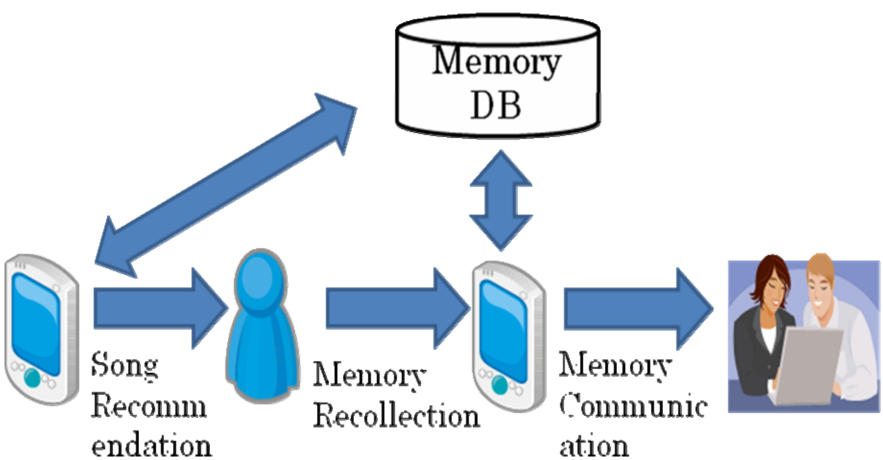


Fig. 2. System configuration

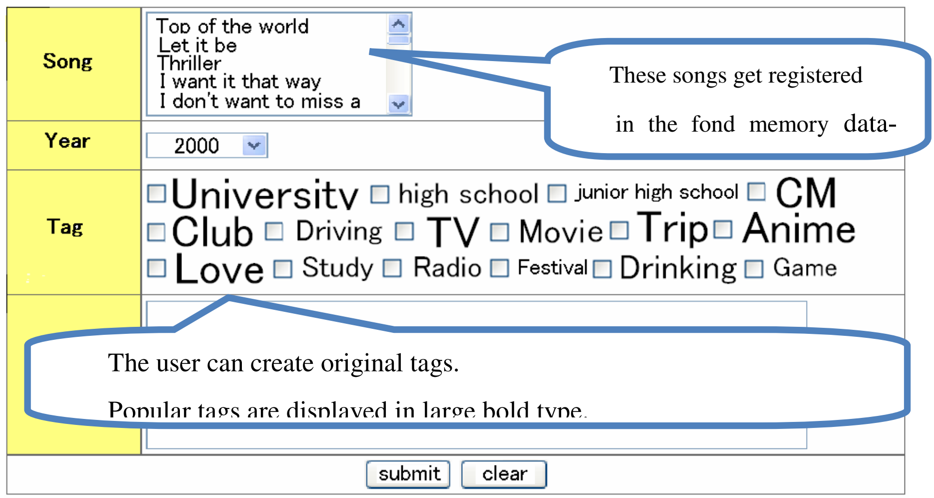


Fig. 3. Screenshot of fond memories being registered

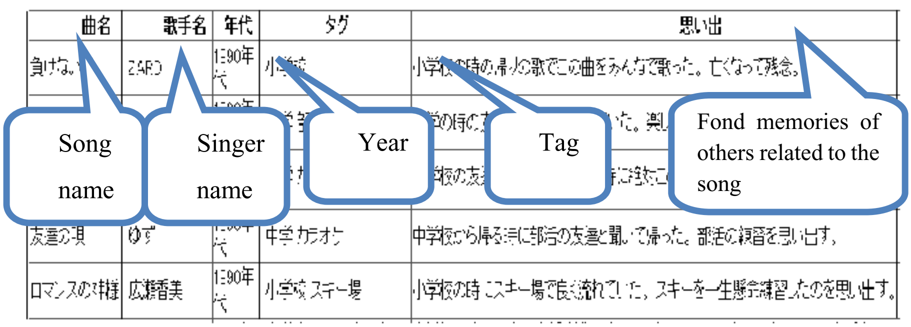


Fig. 4. Screenshot of music recommendation

A system configuration, screenshots of a list of fond memories being registered and music recommendation are shown in Figure 2, Figure 3, and Figure 4, respectively.

The main flow is as follows. First, the system prompts the recollection of a fond memory by providing information on the hummed song or recommending a song through use of rhythm detection. Second the recalled content is stored using a visual image within the system. Finally information on music from users with the same preferences is provided and other people's fond memories then encourage communication.

6 Evaluation

An experiment was conducted to verify effectiveness of this architecture. The participants were 10 Japanese university students, including 8 men and 2 women, varying in age from 21 to 24 years old, with the average age being 22 years old. The participants were asked to use the system in the conversation with the experimenter and to answer the questionnaire after the experiment. The questionnaire consisted of four questions:

(1) Did you recall any memories in use of the system and, if so, what system function prompted it?
  (a) when a name of a hummed song and its singer were identified
  (b) when you looked at the metadata
  (c) when a song was recommended
  (d) when you talked on a recalled memory
(2) Did the system help you have a comfortable talk?
(3) Did you find any new side to your partner?
(4) Did you have feelings of resistance toward the use of this system?

The answers were made in 5-point scale. The result is shown in Figure 5.

When a song was recommended, and when talk was made on a recalled memory, other memories were recalled. In the former case, the song had a link to the other participant's memory. That memory reminded the participants of the similar own memories, which prompted active conversation. In the latter case, different experience about the same song prompted conversation. Both cases helped the participants find new side to their partner. The free answers revealed that conversation about the past experiences prompted deeper mutual understanding, which lead the participants to higher mutual confidence.

One participant had been enjoying quite different music from others of the same age. He felt difficulty to find a song which could be empathized with others. In his case, one-way conversation was observed and conversation was not active. Music recommendation should be refined.

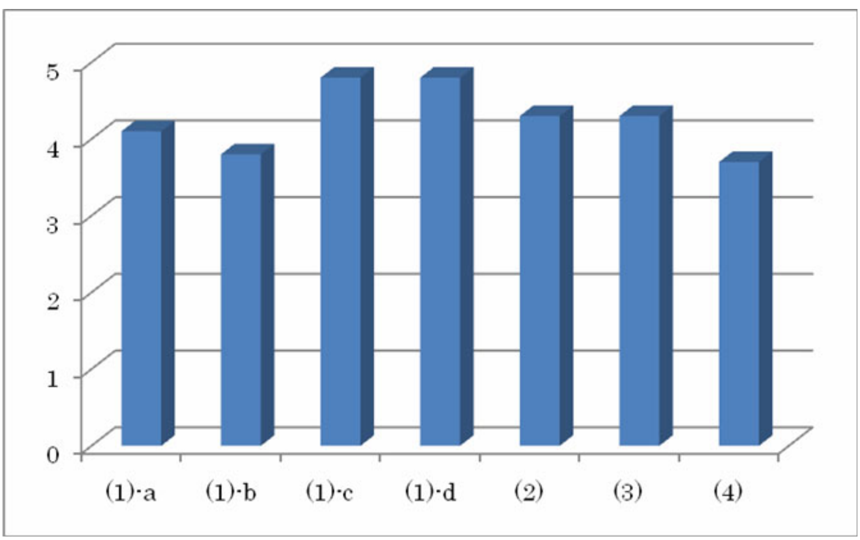


Fig. 5. Answers to the questionnair

7 Conclusion

A recall and communication support system for fond memories triggered by humming was proposed. In the next step, the system is planning to be implemented on the iPhone platform, following which the system will examine how it encourages fond memory recollection and communication.

References

1. Nojima, H., Harada, E.: Cognitive Science in house, Shinyo Company, Capture 12 Memory Engineering, pp. 269–288 (2004)
2. Bluck, S.: Exploring its functions in everyday life. Memory 11, 113–123 (2003)
3. Yamashita, K., Nojima, H.: Portable personal memory for communication. In: Human Interface Symposium (2002) (in Japanese)
4. Nakatani, Y.: Support System for Reconstructing Fond Memories Lost in Disasters. In: Proc. HCI International 2005, pp. 1–10 (2005) (CD-ROM)
5. Munger, D.: Music and memory: How the songs we heard growing up shape the story of our lives. Cognitive Daily (2008)
6. http://scienceblogs.com/cognitivedaily/2008/08/music_and_memory_1.php
7. Melodis Corporation: midomi, http://www.midomi.com/
8. Mathes, A.: Folksonomies – Cooperative Classification and Communication through Shared Metadata, Technical report, University of Illinois Urbana-Champaign (2004)
9. Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., Riedl, J.: GroupLens: An Open Architecture for Collaborative Filtering of Netnews. In: Proc. CSCW 1994, pp. 175–186 (1994)
10. Shardanand, U., Maes, P.: Social Information Filtering: Algorithm for Automating "Word of Mouth". In: Proc. CHI 1995, pp. 210–217 (1995)

Research of Passive Mode Interaction in Pervasive Computing

Yin Lu[1], Kejian Miao[1], Zhanhuai Li[1], and Ke Wei[2]

[1] Computer Science School, Northwestern Polytechnical University
[2] Shaanxi Tourism School
{luyin,kjmiao,zhli}@nwpu.edu.cn,
doreen_weike@hotmail.com

Abstract. Two modes of HCI are discussed in this paper: Active Mode Interaction (AMI) and Passive Mode Interaction (PMI). Closed-loop processing models of each mode are created and analyzed. Contradictory propositions about how to implement spontaneous PMI are identified, and a pipeline model of information transportation in PMI is proposed for a detailed analyzation of mental activity in PMI process. Based on the analyzation, three features of interaction medium that have implications for PMI are identified and principles for PMI interface design are proposed.

Keywords: Human Computer Interaction (HCI); Pervasive Computing; Passive Mode Interaction.

1 Innovation

Artificial Intelligent have greatly changed the way computer interact with human beings. In the previous days, computer operates on human commands, and output results. The human evaluates the outputs with respect to his original intention. If it is not fulfilled, then new directions will be given. It is the human being who takes control of the interaction process. The computer in this process has no difference with common tools without any consciousness. However, in an intelligent computer system, the computer gains the ability of context perceiving and decision making, and works much more conscious then before. It can make its own decision when context changes, and initialize a conversation process to contact the human being, in order to help the user accomplish their goals, or keep away from damages. It is for the first time that machine takes control of interaction process, and try to influence the user's behavior with its output. From the perspective of the human actor, we call the formal mode of human-computer interaction “Active Mode Interaction (AMI)”, and the later “Passive Mode Interaction Process (PMI)”.

The most significant difference between these two modes is the distribution of human attention in the interaction process. In [1], attention is defined as human mental activity that concentrates on a focused target. The concentration of mental activity is a necessary condition of human cognitive activity. The more attention the activity gains, the more efficiency the activity executes. In the AMI process, the user

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 220–229, 2011.

manipulates the whole process, and focus on it with almost all of his attention. While in the PMI process, the user may have been busy with other tasks while interacting with the computers.

The differences between AMI and PMI make some HCI technology developed in the old days work not so efficient in PMI. For example, years ago, we have developed a prototype of a Cooking Assistant pervasive computing service. It tries to help a novice cooker cooking a Chinese meal by providing advices of what spices should be added. It makes use of a set of sensors to trace the progress of the cooking process, and a PDA device to display the recipe and action directions. However, our prototype is rejected by the end user for at least two HCI related reasons:

1. PDA is not a familiar appliance for a cooker. And there is no spare time for a cooker to operate a PDA while busy cooking;
2. Recipes in text format takes time for reading. When the cooker understands what the recipe says, the meal may have burned.

Giving hints in a cooking process is a typical PMI issue. The hints are displayed by the system, and a conversation request is send to the cooker while he is concentrate on cooking. If the interaction request is accepted, the cooker will allocate attention resource for reading hints; if the request is neglected, the interaction process will not be initialized.

Although PDA is objected in our prototype, it is still a powerful HCI medium. It has been used in many applications successfully, e.g. the iHospital project [2], and will be more and more used in pervasive computing services [3]. Now the question is: why doesn't PDA fits for PMI in the Cooking Assistant service? What kind of medium shall be good for that? And how can we make PMI more acceptable by the user? Are there any rules for PMI interface designing?

With the wide spreading of AI, the pervasive computing systems will become more active and smarter than ever[4]. More and more interactions will be carried out in passive mode. If the questions mentioned above are not solved, the awkward interaction process may ruin the "transparency" feature of pervasive computing.

In this paper, we try to figure the features of PMI, and find some principles for PMI interface design. In section 2, we propose closed-loop models of AMI and PMI process, and identify several principles of PMI designing. In section 3, we illustrate four types of PMI process in pervasive computing, and discuss how to implement them; in section 4, we show our implementation of PMI process in the Cooking Assistant service.

2 Analyzation

In *Cognitive Engineering in User Centered System Design*, D.A. Norman proposed a common model of human activities [5]. The model explains how dose a human fulfill his intentions in a closed loop process. However, in Norman's model, the appliances operated by human beings are unconscious tools. The mode is insufficient to describe pervasive computing systems. In figure 1, we propose two closed-loop models in reference of Norman's activity model, to describe AMI and PMI process in pervasive computing.

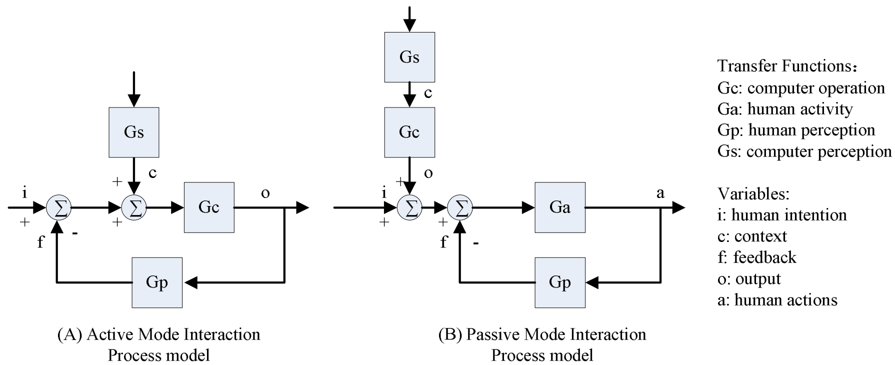


Fig. 1. closed-loop model of AMI and PMI process in Pervasive Computing

In Figure 1: i is the human intention; a is the results of human activity that can be inspected by both the computer and the actor himself; f is the feedback of human activity results, it is the output of human perception; o is output of the computer system that runs on user directions; c is the context information collected by the computer, it is an output of computers perception. Transfer function G_c describes the computing activities manipulated by the user; G_s describe computer's perception that is controlled by AI; G_p describes the inspecting activity of human perception; and G_a describes human activity.

2.1 Why is PMI Process Annoying

In the AMI model, the core of the closed-loop includes user intention, computer operates, user inspection and computer outputs. The sampling object of user perception is the computer outputs. Context information collected by AI is only a reference variable of G_c, and can not be perceived by the human operator directly. While in the PMI model, the core loop only includes activities and perceptions of the human user himself. The computer operates only on base of context information collected by its perceptions. The computer output acts as a forward interference of human intention. If the interference is strong, it will influence the operation of the human user. Otherwise, it may be neglected by the user.

Moreover, in pervasive computing systems, PMI request may interrupt human tasks from time to time. See the timing diagram in Figure 2 part (A). When the pervasive computing system detects a context changing event that is relevant to the human user, it may invoke an interaction request. If the request is accepted by the actor, then he will have to hold on his current task and allocate some attention resources for the coming PMI process. When the interaction is done, the allocated resource will be returned to the actor, and assigned to the user task again. It looks just like the interruption process in computer science. However, human beings have emotions. If their tasks are interrupted too long and too frequently, they may feel annoyed. Then they may neglect the PMI request intensively.

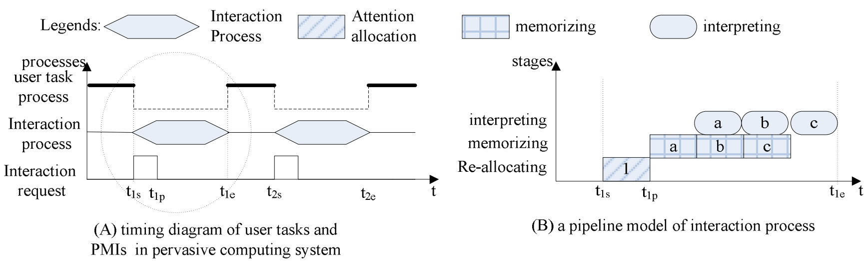


Fig. 2. models of human computer interaction process

2.2 Two Contradictory Propositions

Since the aim of PMI is to help the user with his tasks, or to help him away from potential damages, it is no good to drop a PMI request. The only way to reduce the interference of PMI process is to reduce the attention resources allocated for it. Therefore, there are two contradictory propositions:

1. Firstly, the PMI process should be attractive enough. It has to draw enough attention resources from other concurrent tasks, in order to complete the whole process;
2. Secondly, the total amount of attention resources assigned to it should be limited as little as possible. Otherwise, other concurrent tasks may be hold up by the user because of inadequate resources. More over, if a PMI process attracts too much attention, it may be raised to the current user task, and change into an AMI process.

In the book of Kahneman, he proposes a classic allocation model and some principles of human attention resources distribute among parallel processes [6]. It is said that: Human beings will concentrate more attention on emergent objectives, or something he is interested in. But in the long run, he tends to assign more attention on objectives that he is familiar with, although it may be a cumulative process. When initializing many concurrent processes, the user may estimate the required attention resources of each process. The less the attention required the higher priority the process may be assigned for resource allocation and runtime scheduling.

The model shows that there are different distribution principles in the task initialization and task execution phases. From this point of view, we can divide an interaction process into two phases and solve the contradictory proposition. The former phase is the initialization phase. During this phase, PMI request is accepted and attention resource prepared for the interaction process is allocated. The PMI request should be in the type of something interesting, so as to attract enough attentions, and avoid to be neglected by the user. The latter phase is an information transportation phase. It requires a huge amount of attention resources, in order to interpret information transferred in the interaction.

Although the total amount of attention resources consumed in the transportation phase can not be decreased, the amount of attention allocated from other concurrent tasks can be reduced, because the transportation can be done in a pipeline. We all have the experience of reading. During the reading process, we can not read the whole paragraph at one-time. Instead, we only memorize a short phrase at a glance, and try to understand it while reading the next phrase. The memorizing and interpreting operation iterates till we finished the whole paragraph. In short, the information transportation in reading activity can be described as a two-stage pipeline. The same model fits for information transportation in PMI process either. That is to say, instead of allocate attention resource for information transportation all at once, we can invest only a piece of it, and reuse it in a pipeline in the info transportation phase. The pipeline will reduce the resources required for information transportation significantly.

Figure 2 (B) shows the information transport pipeline model. It is a detailed timing diagram of a complete PMI process, which can be found in the dotted line cycled region in Figure 2(A). The pervasive computing system detects a context event and creates an interaction request at t_{1s}, then the initialization phase starts. In this phase, the user re-allocates his attentions. He draws some resources from the user task, and assigns them to the interaction process. The attention resource is ready at t_{1p}, and then the pipeline of information transportation starts. The user memorizes a piece of information in his mind, and then tries to understand it while memorizing the next piece. This process iterated till the interaction ends at t_{1e}. Because of the pipeline, only a little amount of attention, that is enough to memorize and understand a piece of information, is required at each time during the transportation phase.

In conclusion, interaction request in PMI process must be attractive, and information presentation in PMI should be easy to be memorized and understood. If there is too much information to be presented, it should be divided into pieces to form a transportation pipeline.

2.3 Looking for Interfaces

The discussion above shows how to carry out a PMI process. Now the problem is what kind of interface fits for passive mode interaction. In the previous research of human-computer interaction, many kind HCI technologies have been developed, for example: GUI, context awareness, spontaneous interaction (natural language based interaction, hand writing based interaction and body movements based interaction, etc.). Many of them are developed for information input, such as body movement matching and hand writing recognition. PMI process in pervasive computing is mainly an info output process. If it requires interacting with the operator, it may attract too much attention resources, and change into an AMI process. From the experience gained form Cooking Assistant application, we find that three features should be cared while designing PMI interface: Expressivity, Permanence and Compatibility. The former two features have been discussed in [7]. We will append our point of view gained from pervasive computing system development.

Expressivity: Expressivity relates to the grounding of content [7]. Compared with text, the graphic media is more comprehensive for information expression, and is easier to be understood. Text is more powerful in quantification, conditionals and serialization, while it requires more mental activities for interpreting.

In the two phases of PMI process, graphical media is a better choice of the initialization phase. Interaction request in graphic type is more striking and interesting. It can attract enough attention resource to complete the initialization phase. However, interaction interface for the information transportation should be selected according to the information content. If the amount of information is too much, image that requires one-time attention allocation may consume too may resources.

Video and animation is a kind of enhanced graphical media. It makes up the lack of serialization of graphical media. Vocal or natural language is the reinforcements of text format media. Benefits from the in-directivity of voice, it is much more flexible then pure text format info. However, both video and vocal are dynamic media type, and the lake of permanence limits their utilization in PMI process.

Permanence: The permanence feature is an abstraction of the revisability and reviewability features [7]. Static media type such as image and text can be reviewed at anytime as the user wishes, thus it supports unsynchronized interactions. This feature is important for the initialization phase of PMI process. The PMI process is initialized by an interaction request generated by the computer system. When the request appears, the user may be concentrate on his user tasks. The request must be presented in a static media type, so that it can be accepted at a delayed time.

Compatibility: Compatibility is not the feature of interaction media, but a property of the interface medium. It is a level of accordance between the PMI interface and the running user task.

The research of common sense has revealed that each kind of user task has a rather static array of activities and a collection of appliances [8]. If the PMI medium has something related to a member in the appliance collection, we can say that the medium is compatible with the user task. If the medium is included in the collection, then we say that they are fully compatible. A medium with compatibility will be more acceptable by the user in the PMI process.

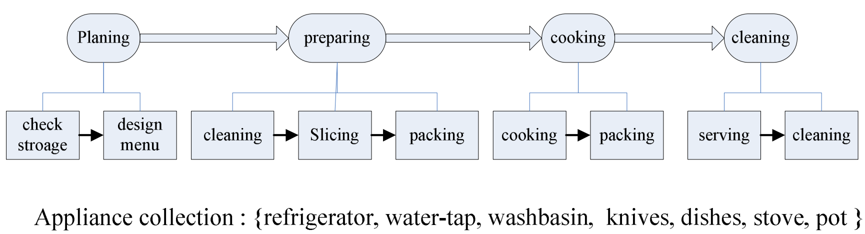


Fig. 3. Activity array and appliance collection of a cooking task

For example, Figure 3 shows the activity array and appliance collection of cooking a meal. From the experience we gained in smart home application developing, we know that if we take a PDA as PMI medium device in a cooking task, the end user will object our system. However, if we take a condiment bottle as a medium to show hints to the cooker, the user will accepte it. The condiment bottle is appliance that compatible with the cooking task. Because it has a property that is the same as the member appears in the applicance collection: in the kitchen.

2.4 PMI Interface Based on Perceived Affordance

The concept of "Object Affordance" is proposed in [9] to refer to the actionable properties between an object and a human actor. D.A Norman discussed it in [10], and proposed "Perceived Affordance" in [11]. Perceived affordance is about the actions a user perceives to be possible, it may not the actions that the object really has. Perceived affordance is influenced by the knowledge and experiences of the actor.

Members in the appliances collection of a user task have strong properties of "Perceived Affordance". When the human user sees a member object, he may think of activities in the activity array all at once. For example, when a patient sees his drug box, he may think of taking medicine; when a cooker sees a kitchen knife, he may think of slicing meets and vegetables into pieces.

Perceived affordance is useful in passive mode human-computer interactions. Objects with perceived affordance property can be used as an interaction medium. It works like a kind of enhanced graphical interface. The media of the medium is its shape, color, or other object properties, and the information content is its perceived affordance. Compared with other kind of medium, a perceived affordance medium is fully compatible with the user task in nature.

3 PMI in Pervasive Computing

From the previous discussion, we know that in order to implement a successful PMI process, we need a brilliant interaction request, an efficient way of information transportation and a compatible interaction medium. These are only principles for PMI designing, and the detailed implementation depends on the information transported in the process.

Based on research of smart home pervasive computing systems, we sort out four types of PMI process in pervasive computing: activity reminds, status reminds, information service and emergency alert. In this section, we will discuss the features of them, and give some proposals for the implementation.

Activity Reminds: In an activity reminds PMI process, the pervasive computing system tries to remind the actor of some activities that he may forget. The interaction request of it may be emergent, but the amount of information transferred is tiny. An abstract icon image or several words about the activity is enough to remind the actor about it.

Interface based on perceived affordance is another good choice. We can notify the user with flashes or sounds. And as soon as he sees the medium object, the affordance property may tell him what to do.

Status reminds: In a status reminds PMI process, the computing system tries to show the running status of the user task, in order to help the user restore an interrupted task. The interaction request is not as emergent as in activity reminds, but the amount of information transferred may be greater. Graphic or text based interface are all acceptable in this type of PMI process.

Information Services: Information service is a kind of agent program. The pervasive computing system collects and refines information on behalf of the user, and presents it in a way that is easy for reading. PMI process in information service is not emergent. It even does not need an interaction request. But the information presented must be permanent, and can be visited at any time the user wants. The interaction medium should be compatible with the user living habits. For example, if the user used to see if it is raining when he tends to go outside, then the detailed weather forecast information should be displayed on a screen beside the door, or directly projected on the window in his house.

Emergency alert: Emergency alert is not a pure PMI process. It requires emergent interacting with the user. The PMI process will shift into AMI process when the user accepts the interaction request. A GUI on a embedded device, or natural language based interface will be good in this type of PMI.

4 Implementation

Based on the discussion above, we rebuilt the Smart Cooking Assistant context-aware Service (SCAS) mentioned in section 1.

The SCAS maintains a knowledge database of Chinese meal recipes. The recipes are collected from internet automatically. Before cooking a meal, a menu should be input to the service. This is an AMI process, and can be completed on a desktop computer, or on a smart phone.

SCAS inspects the cooking process by a set of pressure sensor. The sensor is attached to the bottom of each condiment bottle. When the cooker begins to deal with the meal, SCAS will tell the cooker what kind of spice should be added, and how many spices he has added. This is a status reminds PMI process. Information delivered in the process shows the status of cooking a dish.

SCAS takes condiment bottle as the interaction medium. We have attached two led lights to each condiment bottle, one is group light, and the other is status light. The group light on means that this kind of spice is included in the recipe of the dish. The status light shows whether the spice has been added to the dish. In short, spice in a condiment bottle with both group light and status light on, should be added to the dish.

A detailed description of the implementation of SCAS can be found in [12].

5 Related Works

In recent years, many progresses have been made in HCI research. Among them, the most important achievement for pervasive computing is multi-modal interaction. Since human interactions are multi-modal in nature, interaction in pervasive computing can not be restricted within desktop anymore. It requires spontaneous interaction in multi-channels.

The topic of multi-modal interaction has been discussed over and over in the region of artificial intelligence. In the series work of Steve Whittaker's group, they evaluated varied communication channels and proposed a set of principles of multi-modal interaction [13,14,7]. AAAI symposiums [15,16,17] discussed the relationship and interaction between human user and intelligent systems. The achievements made in AI research improve the interactivity of pervasive computing systems, and implement implicit interaction channels with explicit GUI based interaction channels [18]. Many experimental systems have been built in recent years, and interaction models were proposed corresponding to the systems, e.g. [19,20].

Besides, interactions on limited size screen are also a hot spot in HCI research, on account of the repaid progress of mobile computing and ad hoc networks. In paper [21,22] the authors developed pervasive applications running on smart phones. And in [3], mobile network terminals such as smart phones are considered as an ideal interacting medium for pervasive computing.

Since that HCI is a communication process between computer and human user, improvements made in technical region is inadequate to implement successful spontaneous HCI. Discussions about human nature are also necessary. AAAI symposium [23] discussed the issue about initiative in HCI. In paper [24], the authors discussed the mental activity in pervasive computing HCI, and propose a multi-modal interaction framework. Paper [25] reviewed the development of human-computer interaction technology, and discussed the influence of culture and habits.

Most of the research mentioned above, focus on providing methods and devices that facilitating the accession of a cyber world maintained by the computers. However, not all of the tasks in our everyday life can be accomplished in the cyber world. How to communicate with the human actors when they are busy doing these tasks, is an issue seldom discussed previously. In this paper, we propose the concept of Passive Mode Interaction, and try to answer the question from our point of view.

References

1. Li, W.: Research on Congnitive Psychology, pp. 103–105. Zhejiang People Press, Hangzhou (1998) (in Chinese)
2. Thomas, R.H., Jakob, E.B., et al.: Moving Out of the Lab: Deploying Pervasive Technologies in a Hospital. IEEE Pervasive Computing 5(3), 24–31 (2006)
3. Rafael, B., Jan, B., et al.: The Smart Phone: A Ubiquitous Input Device. IEEE Pervasive Computing 5(1), 70–77 (2006)
4. Juan, C.A., Chris, D.N.: Smart Homes Can Be Smarter. In: Smart Homes: The Role of Artificial Intelligence, 2nd edn. Springer, Berlin (2006)
5. Donald, A.N.: Cognitive Engineering in User Centered System Design. Lawrence Erlbaum Associates, Hillsdale (1986)
6. Kahneman, D.: Attention and Effort. Prentice Hall, New York (1973)
7. Steve, W., Marilyn, A.W.: Toward a Theory of Multi-Modal Interaction. In: AAAI Workshop 1991, pp. 78–85. AAAI Press, Menlo Park (1991)
8. Liu, H., Singh, P.: Commonsense Reasoning in and Over Natural Language. In: Negoita, M.G., Howlett, R.J., Jain, L.C. (eds.) KES 2004. LNCS (LNAI), vol. 3215, pp. 293–306. Springer, Heidelberg (2004)
9. Gibson, J.J.: The Theory of Affordances. Perceiving, Acting, and Knowing. and Knowing. Lawrence Erlbaum Associates, Hillsdale (1977)
10. Donald, A.N.: The Psychology of Everyday Things. Basic Books, New York (1998)
11. Donald, A.N.: Affordance, Conventions, and Design. ACM New York Interaction 6(3), 38–43 (1999)
12. Yin, L., Kejian, M., et al.: Cooking Assistant Service in Smart Home. Computer Engineering and Applications 45(36), 189–192 (2009)
13. Steve, W., Susan, B., Herb, C.: Coordinating Activity: an Analysis of Interaction in Computer Supported Cooperative Work. In: ACM SIGCHI 1991, pp. 1–9. ACM Portal, New York (1991)

14. Steve, W.: Simulating Agency. Technical report, Hwelett Packar Laboratories, Bristol, England (1990)
15. David, K., Michael, F.: Human Interaction with Autonomous Systems in Complex Environments. In: AAAI Symposium 2003 (SS-03-04). AAAI Press, Menlo Park (2003)
16. James, G., Cheryl, M.: Interaction between Humans and Autonomous Systems over Extended Operation. In: AAAI Symposium 2004 (SS-04-03). AAAI Press, Menlo Park (2004)
17. Neil, Y.S.: Interaction Challenges for Artificial Assistants. In: AAAI Symposium 2007(SS-07-04). AAAI Press, Menlo Park (2007)
18. Kamel, N., et al.: A Formal Model for CARE Usability Properties Verication in Multimodal HCI. In: IEEE SEPS 2007 Workshop (joint with MAPS 2007 Workshop), July 2007, pp. 341–348. IEEE Press, New York (2007)
19. Horchani, D., Fréard, E., Nigay, L., et al.: A Platform for Designing Multimodal Dialogic and Presentation Strategies. In: 11th Workshop on the Semantics and Pragmatics of Dialogue, DECALOG 2007, Trento, Italy, vol. 2007, pp. 169–170 (2007)
20. Asare, P.: A Sign of the Times: A Composite Input Device for Human-Computer Interactions. IEEE Potentials 29(2), 9–13 (2010)
21. Jeffrey, N., Brad, A.M.: Controlling Home and Office Appliances with Smart Phones. IEEE Pervasive Computing 5(3), 60–67 (2006)
22. Adrian, D.C., Anuroop, S., et al.: Capture the Flag: Mixed-Reality Social Gaming with Smart Phones. IEEE Pervasive Computing 5(2), 62–69 (2006)
23. Susan, H., Susan, M.: Computational Models for Mixed Initiative Interactions. In: AAAI Symposium 1997(SS-97-04). AAAI Press, Menlo Park (1997)
24. Wei-ning, Y., Shi-hai, D., et al.: Study on Human-Computer Interaction Framework of Pervasive Computing. Chinese Journal of Computers 27(12), 1657–1664 (2004) (in Chinese)
25. Jonathan, G.: Three Faces of Human-Computer Interaction. IEEE Annals of the History of Computing 27, 46–62 (2005)

Activity Recognition for Risk Management with Installed Sensor in Smart and Cell Phone

Daisuke Honda, Nobuchika Sakata, and Shogo Nishida

Division of Systems Science and Applied Informatics Graduate School of Engineering Science, Osaka University 1-3, Machikaneyama-cho, Toyonaka-city, Osaka, 560-8531 Japan
{suzuki,sakata,nishida}@sys.es.osaka-u.ac.jp

Abstract. Smart and cell phone with self-contained sensor such as accelerometer, gyroscopic and digital magnetic compass sensor have been popular. Combining certain algorithm and those sensors, it can estimate user's activity, situation and even user's absolute position. However, estimation of user's activity, situation and user's absolute position become difficult when once sensors posture and position are changing from original position in user's motion. Also, according to stored, worn and handheld position and posture of those cell and smart phone are often changed. Therefore, we exclude estimation of user's position and we focus to only estimation of user's activity and situation for risk management. Basically, we design special classifier for detecting user's unusual behavior and apply other user's position data from internet to the results detected by the classifier which are combined wavelet transform and SVM. We assume that user's unusual activity and situation can be detected by smart and cell phone with high accuracy.

Keywords: Activity recognition, Wearable computer, SVM, nearly fall incident, cell phone and smart phone.

1 Introduction

In 1929, Heinrich a safety engineer in U.S. found the principle that there are 29 small accidents and 300 incidents behind 1 fatal accident. This principle tells us the way to prevent fatal accidents. To prevent fatal accidents, we should not overlook small accidents and incidents and should take countermeasure based on it. Aviation industry in U.S. founded ASRS in 1976 and started incident report system. Learning from mistake is more profitable than punishment to mistake they thought. Also they have achieved a certain result. Currently, the importance of incident report is acknowledged in all over the world. There are a lot of organizations conducting incident report system, such as IMO, MAIB, MARS, IMISS and TSB.

Community involving high-risk job specialist such as aviation industry, medical field and marine tend to be adopted incident report activity. On the other hands, in community involving low-risk, it is difficult to take such kind of incident report activity. People cannot keep motivation and perceive incentive to report their incident experiences. Finally, such incident report system is adopted only after a fatal accident happens.

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 230–239, 2011.

For this problem, automatic accumulation system of incident information has been tried in motor vehicle industry. InterNavi[1] developed by HONDA is actual system of road maintenance and improvement using information of sudden braking cases accumulated automatically from general public drivers. Conducting road maintenances based on danger place and point obtained from accumulated data involving sudden brakes, and then, the number of sudden braking was reduced 70%. Currently, such activity is not tried to pedestrian. Therefore, we assume that there is effect to reduce the fatal incidents with applying automatic incident accumulation to pedestrian. In order to apply automatic incident accumulation system to pedestrian, we consider major two problems not to be found in motor vehicle. First, there are a lot of flexibilities in human movement when comparing to motor vehicles. There is no obvious key to detect incident occurrence, such as sudden braking in motor vehicle. There are many researches for studying detection of complete fall accidents [2],[3],[4]. The detection key of complete fall accidents is clearer than that of nearly fall incident. Second, there is a practical problem to deploy sensors on pedestrian's body.

2 Policy of Sensor Deployment

Motor vehicle has been computerized drastically. There are a lot of ICs and sensors everywhere in a modern car. In case that we can deploy ICs and sensors on human body wherever we like, human activities can be recognized automatically to some extent. However, it is difficult to deploy sensors on human body wherever we like. In case of guarantying a lot of merits directly for the person to wear computers and sensors, pedestrians may wear computers and sensors. For example, smart computer realize context awareness [5], learn application for commuter adapting to user context [6], and monitor health care of elderly people [7] have been reported. In such applications, single purpose sensors are deployed at upper arms, thighs, shoes and cane wherever and whatever developer like. However, there is no merit for each pedestrian directly in the application we propose. The merit of our proposed application is not for individual but for community. Without profiting individually merit, people may not accept burden of wearing special sensors. In addition, incidents of pedestrian are not occurs frequently. Thus single purpose device is not worth for our proposed application in terms of time cost, wearable burden and bothers.

Considering constraint above-mentioned, the most practical way is to use mobile phone. Almost of all pedestrian carry mobile phone. Also smart phones equipped with inertial sensor are quite popular such as iPhone. We assume that more and more people may have it in the future. With using installed sensors in mobile phone, we can induce many pedestrian to accept automatic incident report system with no concern of user burden and cost of special device.

When we use the sensor in mobile phone, we must consider the stored position of mobile phone on human body. iShare Corporation reported one of the result of awareness survey [8] about mobile phone. According to this survey, stored position of mobile phone varies from person to person, and also varies from day to day in same person depending on attire or activity. In case building applications of human activity recognition with installed sensors in mobile phone, recognition algorithms should be changed according to the stored positions of mobile phone. Basically, device contexts should be recognized first. After that user context can be recognized. We assume that Two-phased approach is needed.

Fujinami et al. conducted a research about device context of mobile phone with tri-axial acceleration sensor [9]. According to their research, stored positions of mobile phone can be recognized with appropriate accuracy using sensor signals of stationary movement period like walking. Therefore, we suppose that the stored position of mobile phone is obtained from their method. In this paper, we only focus a method to detect nearly fall incidents occurrence during walking.

Applications for individual user need perfectible recognition of user activities even in any stored position and posture of mobile phone. However, our proposed application is for community. We assume that the automatic incident reports system gather a lot of data from many people carrying with cell and smart phone. Thus, our proposed application does not chase individual activities perfectly. The application can ignore difficult device context with a lot of flexibility such as handheld situation or holding in bag. Finally, we plan to activate incident detection application only when mobile phone is in bottom of user's pocket. Bottom of user's pocket can be expected not to have considerable individual difference in shape. At the same time, bottom of user's pocket is close to center of gravity of human body. It can be expected that discriminative signal is observed when the person lose balance by trip and slip. In addition, it is possible to apply to one-third of pedestrians even if we ignore the position other than bottoms of user's pocket from ishare report result [8]. In the future, there is possibility to extend the scope of application to other stored position of mobile phone. However, in this paper, we believe that it is enough for initial approach.

3 Detection Method

In this research, we use Wii and Wii motion plus instead of smart phone equipped with sensors to measure human activity. It contains a tri-axial acceleration sensor and a tri-axial gyro sensor. We configure sampling frequency at 16 Hz. It is determined consulting with the research [10], [11]. Also we suppose that device context is in bottoms of user’s pocket (Fig. 1). There are three chief functions.

Fig. 1. Appearance of sensor position. We plan to activate incident detection application only when mobile phone is in bottom of user's pocket. Bottom of user's pocket can be expected not to have considerable individual difference in shape.

1. Transformation into coordinate independent of sensor attitude.
2. Division of time-series data and feature extraction.
3. Learning and recognition by SVM.

3.1 Transformation into Coordinate Independent of Sensor Attitude

We suppose that mobile phone is in bottoms of user's pocket. Even so, there is flexibility of sensor attitude. We can simplify and exclude the flexibility by the following steps. In order to exclude the flexibility of sensor attitude, we transform the original sensor data into the data according to coordinate on pedestrian. (Fig. 2) The three axis are consist of

1. Direction of gravitational force,
2. Direction of pedestrian's movement
3. Side direction of pedestrian.

Transformation of coordinate is conducted during a stationary period of walking. Mobile phone in the bottoms of user's pocket is in certain restraint. It is unthinkable that mobile phone spin round and round in the bottoms of user's pocket. Small rotary motion may be occurred along to walking movement, but it can be expected like the swing of pendulum. Also device may keep stable attitude in average viewpoint.

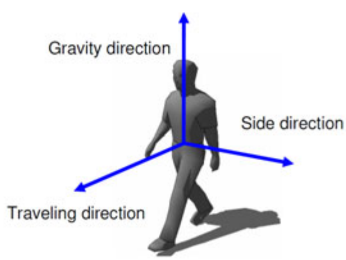


Fig. 2. Coordination on pedestrian. We transform the original sensor data into the data according to coordinate on pedestrian.

For this reason, we simply evaluate gravity acceleration vector by putting raw data of tri-axial acceleration sensor through low pass filter. We use third degree Butter worth low pass filter. Cutoff-frequency of the filter is configured at 0.5Hz. Signals under 0.5Hz is unlikely to be mixed in, no matter how slow the walking pace is. We can obtain nearly flat signal by this way. Also it is the approximation of gravity acceleration vector. (Fig. 3)

Next, we exclude element parallel to the gravity acceleration vector from the signals of tri axial acceleration sensor. We obtain the acceleration element existing in horizontal plane. In horizontal plane, swing along travelling direction is larger than that alongside direction. We can obtain the vector of travelling direction axis using principle component analysis (PCA). The vector of side direction axis can be obtained from pre mentioned two axis. By this way, we can obtain three axis installed on pedestrian's body. Along to the three axes, we can obtain acceleration and gyro signals independent of sensor attitude. (Fig. 4) The process above mentioned uses duration of stationary walking activity for 2 second. This process is based on the assumption that mobile phone attitude is remained steadily in average viewpoint. It cannot be used under the condition having more flexibility. However, it is enough for the condition of bottoms of user's pocket.

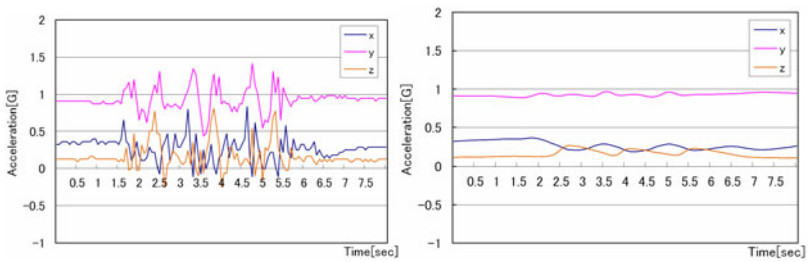


Fig. 3. We use third degree Butter worth low pass filter. Cutoff-frequency of the filter is configured at 0.5Hz. Left side figure is raw data from sensor. Right side figure is after applying the low pass filter.

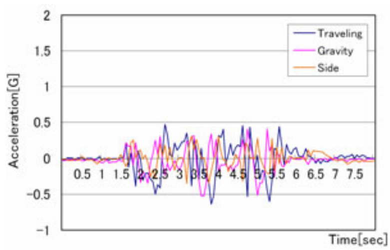


Fig. 4. The vector of side direction axis can be obtained from pre mentioned two axis. By this way, we can obtain three axis installed on pedestrian's body. This figure shows sensor data removed gravity.

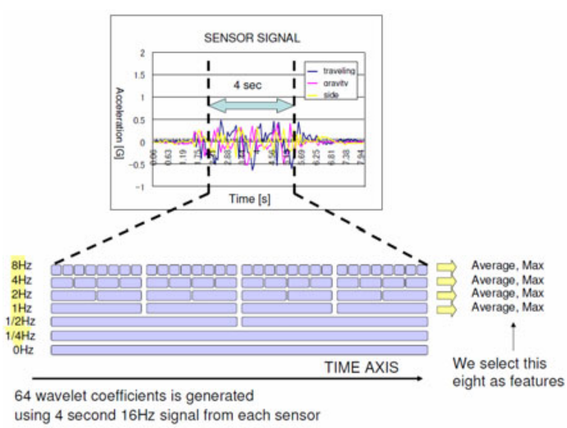


Fig. 5. We calculate maximums and averages of 1Hz, 2Hz, 4Hz and 8Hz coefficients individually. By this way, we obtain 8 features for each sensor. In this phase, time flexibility disappears.

3.2 Division of Time-Series Data and Feature Extraction

Pattern recognitions of time-series data require the process of dividing the data and feature extraction from the divided data. This time, we divide sensor data at the interval of four second. There are six sensors. Sampling frequency is 16 Hz. Thus, one compartment consist of $4 \times 6 \times 16 = 384$ data. In case of that we use this 386 data as a feature vector, there remains flexibility of time axis, and the flexibility make pattern recognition process complex. In order to exclude flexibility, feature extraction is generally conducted before pattern recognition process. In speech recognition, transformation into frequency domain is commonly used such as Fourier transformation. However, signal of nearly fall incident is not a periodic signal. It is sudden signal like single peaked pattern. We assume Fourier transformation that intend for periodic signal is not suitable, and conduct feature extraction by Wavelet transformation which is strong for the analysis of sudden signals. In this time, we divide signals simply into Haar wavelet. We apply wavelet transformation to each sensor data series. Applying wavelet transformation to 64 data (16Hz × 4sec), we obtain the same number of wavelet coefficients which consist of one coefficient corresponding to 0Hz, one coefficient corresponding to 1/4Hz, two coefficients corresponding to 1/2Hz, four coefficients corresponding to 1Hz, eight coefficients corresponding to 2Hz, sixteen coefficients corresponding to 4Hz, thirty two coefficients corresponding to 8Hz. Using wavelet transformation, there remain time flexibilities in this phase. We calculate maximums and averages of 1Hz, 2Hz, 4Hz and 8Hz coefficients individually. (Fig. 5) By this way, we obtain 8 features for each sensor. In this phase, time flexibility disappears. There are six sensors. So now we obtain 48features. We add variances of each sensor data and finally obtain 54 features in total.

At the last step, we normalize 54 features individually. For this normalization, we calculate averages and variances of each feature from data of various activities of various people.

3.3 Learning and Recognition by SVM

SVM is often-used algorithm of pattern recognition because of its high performance in experimentation, existence of theoretical basis, existence of simple implementation, and capability of nonlinear recognition using Kernel method [12].

We use the SVM algorithm written in Reference [13]. We use Gaussian Kernel. The parameter variance of this Kernel is configured at 2000. Another parameter of soft margin is configured at 1000.

4 Experiment

Our research objective is detecting nearly fall incidents of pedestrian in outdoor situation of regular daily activities. In other word, we suppose to distinguish nearly fall incidents from daily activities. The method described above is constructed for this purpose. Currently, we must evaluate classification ability of the method. In order to evaluate the method, we need to collect samples of nearly fall incident and regular daily activity. Collecting real incident data sufficiently in real life is very difficult, because incident do not occur frequently. Therefore we collect incident samples by imitation of subject. Also we define four activities as regular daily activities at outdoor. The four activities are "WALK","RUN","GO UPSTAIRS", and "GO DOWNSTAIRS". We let subject perform the four activities in small excursion. Then we define incidents as nearly fall incidents this time. The incidents can be caused by each side of foot, by right foot or left foot. By contrast mobile phone is held in only one side of the bottoms of user's pocket. Thus, we define following two incidents; "SENSORSIDE TRIP" and "OTHER SIDE TRIP".

Resultantly, we collect samples under six type conditions, "WALK","RUN","GO UPSTAIRS", and "GO DOWNSTAIRS" as daily activity "SENSOR SIDE TRIP" and "OTHER SIDETRIP" as nearly fall incident.

We let twelve subjects (man, age 22-25) serve each activity ten times. As a result, we collect total 720 samples.

Fig. 6 shows the tendency of 54 features distribution in "SENSOR SIDE TRIP". Unfortunately we cannot illustrate the 54-dimenal space. We compromise and illustrate the tendency of distribution by box-and-whisker plot.

As a pilot test, we divide 720 data into two group consist of 360 data. Both groups contain each activity of each subject in same number. First, we use one group as learning and conduct recognition test to other group. Then, we exchange the group role and conduct the same test and the result is shown in Fig. 7. There are 120 samples in each activity. Also figures in the table are number of successful samples of recognition. Frequent false positive is fatal obstacle to the incident. So smallness of false positives most important criterion we consider. It is 94.2%. To achieve it near 100%, additional improvement is needed. Evaluating it in often used criterion, the result is 88.0%Precision, and 85.8% Recall.

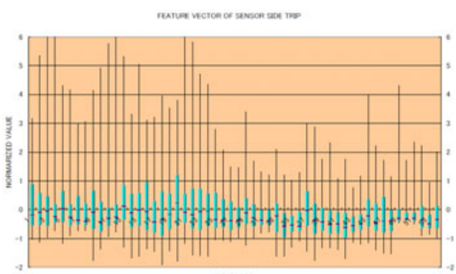


Fig. 6. Figure shows the tendency of 54 features distribution in "SENSOR SIDE TRIP"

USING 1-54 (ALL FEATURES)

	success	%		success	%
WALK	116	96.7	*DAILY*	452	94.2
RUN	116	96.7			
GO DOWNSTAIRS	107	89.2			
GO UPSTAIRS	113	94.2			
SENSOR SIDE TRIP	98	81.7	*INCIDENT*	206	85.8
OTHER SIDE TRIP	108	90.0			

Fig. 7. We use one group as learning and conduct recognition test to other group. Then, we exchange the group role and conduct the same test and the result.

5 Discussion

Using 54 features, we attain 85.8% recall with false positive 5.8%. And precision is 88.0%. Actually this precision cannot be applied to real life data. Generally speaking, nearly fall incident is slightly rare. Even if false positive rate is quite low level, false detection can be accumulated in long daily activities. Consequently, precision is decreased in real life situation. Therefore, more effort and improvement is needed.

On the other hand, automatic incident report system does not need complete accuracy. It is enough to discover the candidate sites having a potential danger. With applying the application to sufficient number of pedestrian, the number of detected incidents at dangerous site may be expected to grow larger comparing with the number of false detections at not dangerous site.

Table 1. The result of recognition with smaller size feature vector

USING FEATURES	PRECISION %	RECALL %
ALL FEATURES	88.0	85.8
ALL WAVELET FEATURES	87.2	85.0
VARIANT OF RAW SIGNAL FEATURES	36.1	88.3
8HZ WAVELET FEATURES	84.0	67.9
4HZ WAVELET FEATURES	86.3	68.3
2HZ WAVELET FEATURES	78.2	70.4
1HZ WAVELET FEATURES	76.6	54.6
MAX OF WAVELET FEATURES	63.9	53.8
AVERAGE OF WAVELET FEATURES	78.2	65.8
ACCELERATION FEATURES	81.6	66.7
GYRO FEATURES	85.3	75.0

Next, we conduct learning and recognition by SVM with smaller size feature vector to see contributions to the recognition of each feature. Table 1 show the result of recognition with smaller size feature vector. As shown in Table 1, we can realize the fact that VARIANT OF RAW SIGNALFEATURES does not contribute so much. The low precision in the line of VARIANTOF RAW SIGNAL FEATURES and high precision and recall in the line of ALLWAVELET FEATURES show the fact. Also you can see the fact that contribution of WAVELET FEATURES makes little difference by changes in frequency. Next you can see the fact that AVERAGE OF WAVELET FEATURES makes larger contribution compared to MAX OF WAVELET FEATURES. Previously, we use only MAXOF WAVELET FEATURES, regarding it as best to detect sharp peak due to tripping and losing balance. Precision is 63.9% and recall is 53.8%. Combining AVERAGE OFWAVELET FEATURES, precision improved 23.3% and recall improved 31.2%. More preferable selection may be existing in 64 wavelet coefficients corresponding 4 second of one sensor. GYRO FEATURES make larger contribution compared to ACCELERATIONS. Using acceleration sensor only, we obtain 81.6% precision and 66.7% recall. Combination use of gyro sensor improves the precision by 6.4% and the recall by 19.1%.For this reason, gyro sensor is essential to high accuracy. Recognition by gyro sensor only cannot work in our framework. Because of that sensor context recognition by Fujinamiet.al uses acceleration sensor, and detection of gravity direction also needs acceleration sensor.

In conclusion, all of 54 features are contributing to a greater or lesser extent. However VARIANT OF RAW SIGNAL FEATURES make less than 1% improve both in precision and recall. Other features are all essential to high precision and recall. For more improvement, incorporation of another sensor data or incorporation of some kind information or twist in feature extraction are hopeful.

6 Conclusion

In this research, we tried to implement nearly fall incident detection system, assuming that mobile phone is in bottom's pocket of pedestrian. We tested the system by imitated trip incident sample and daily activities. As a result, we obtained 88.0% precision and 85.8% recall (and False positive rate is 5.8%). We confirm the delectability of nearly fall incident by imitated samples at first step this time.

Acknowledgments. This research was partially supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Young Scientists (B) "Remote collaboration with Real-world based Interface", No.20700111.

References

1. http://www.honda.co.jp/internavi/ (in Japanese)
2. Lan, M., Nahapetian, A., Vahdatpour, A., Au, L., Kaiser, W., Sarrafzadh, M.: SmartFall:An Automatic Fall Detection System Based on Subsequence Matching for the SmartCane. In: BodyNets 2009 (2009)
3. Yoshimura, T., Nakajima, K., Tamura, T.: An ambulatory fall monitoring system for the elderly. In: Third International Conference of Gerontechnology, Munich, Germany, pp. 10–13 (October 1999)
4. Shinagawa, Y., Tanigawa, T., Ohta, S.: Detection of Human Gait and Falling Using an Accelerometer. Kawasaki Journal of Medical Welfare 9(2), 243–250 (1999) (in Japanese)
5. Takada, D., Ogawa, T., Kiyokawa, K., Takemura, H.: A Context-Aware Wearable AR System with Dynamic Information Detail Control based on Body Motion. Journal of Human Interface Society 12(1), 47–56 (2010) (in Japanese)
6. Nakamura, T., Ogawa, T., Kiyokawa, K., Takemura, H.: Implementation and Evaluation of a User Context Recognition Mechanism for Continuous Wearable Learning Systems. Technical Report of IEICE 106(396), 17–22 (2006)
7. Lan, M., Nahapetian, A., Vahdatpour, A., Au, L., Kaiser, W., Sarrafzadh, M.: SmartFall:An Automatic Fall Detection System Based on Subsequence Matching for the SmartCane. In: BodyNets 2009 (2009)
8. iShare Inc., Survey on mobile phone usage (November 2005), http://bizmakoto.jp/bizmobile/articles/0511/04/news105.html (in Japanese)
9. Fujinami, K., Jin, C., Kouchi, S.: Tracking On-body Location of a Mobile Phone, Late Breaking Results - Cutting Edge Technologies on Wearable Computing. In: The 14th Annual IEEE International Symposium on Wearable Computers, ISWC 2010 (October2010)
10. Bouten, C.V.C., Koekkoek, K.T.M., Verduin, M., Kodde, R., Janssen, J.D.: A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity

11. Tabuchi, K., Naya, F., Ohmura, R., Noma, H., Kiyoshi, K., Kishino, F.: Evaluation of data collection parameters on the daily activity classification with accelerometers. IEICE Technical Report PRMU, 43–48 (May 19,2006) (in Japanese)
12. Tsuda, K.: Overview of Support Vector Machine. Technical Survey IEICE 83(6), 460–466 (2000) (in Japanese)
13. Introduction to Support Vector Machine (in Japanese) ISBN-13: 978-4320121348

Can Twitter Be an Alternative of Real-World Sensors?

Tetsuro Takahashi, Shuya Abe, and Nobuyuki Igata

Fujitsu Laboratories, Ltd.
1-1, Kamikodanaka 4-chome, Nakahara-ku, Kawasaki 211-8588, Japan
{takahashi.tet,abe.shuya,igata}@jp.fujitsu.com

Abstract. Twitter is the most famous on-line microblogging service now. People can post (tweet) what they are doing in 140 characters. Since Twitter posts (tweets) reflect what people are looking, hearing, feeling and so on, we can obtain information about Real-world phenomena through the large amount of tweets. In other words, Twitter can be regarded as a sensor of Real-world phenomena including natural phenomena such as hay fever. This motivated us to investigate whether can Twitter be an alternative of Real-world Sensor. In this paper, we first describe about our system which collects and analyzes tweets in order to generates a hay fever map just like as a weather report map. There are some difficulties such as location estimation and normalization of number of tweets. Using the output of the system, we discuss the comparison with actual pollen data gathered by real sensors. The result shows that Twitter can reflect natural phenomena in some particular areas.

Keywords: Twitter, crowd knowledge, social sensor.

1 Introduction

Twitter is the most famous online microblogging service. People can post messages but are restricted to 140 characters.[1] Since tweets reflect what people are doing, seeing, hearing and feeling, we can obtain information about real-world phenomena through the tweets. In other words, Twitter can be regarded as a sensor of real-world phenomena, including natural phenomena, such as hay fever. Once we prove the possibility of Twitter as a sensor for real-world phenomena, we will be able to measure various phenomena without actual sensors. Furthermore, we can also acquire immeasurable data, such as how happy or tired people might be. This motivated us to investigate whether Twitter could be an alternative to real-world sensor. Based on this background, we chose hay fever (pollinosis) as the target of our experiment. This target is convenient for evaluation, because we could obtain actual data from real-world sensors for pollen.

Twitter is categorized as Consumer Generated Media (CGM). The most typical CGM is blogs for which there are many services and a great deal of research has been done [1]. Compared with blogs, Twitter has at least two major characteristics:

[1] We call the posted text a “tweet” instead of another name like status, messages or posts.

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 240–249, 2011.

- *Real-time nature*
  While blogs are posted one to several times a day, tweets are posted many times a day. People post "what is happening now" in real-time rather than as messages posted each day.
- *Geolocation information*
  Twitter can attach geolocation information (latitude and longitude) to each tweet. People often use mobile devices to tweet compared to blogs, because of the restriction on text length and the nature of the service where people input their status.

These characteristics tell us that Twitter is more appropriate than the former CGM as the information source for sensors.

There are many researches on Twitter that have examined Twitter from various aspects, such as usage [2], social network [3] and communication [4]. Regarding the idea of using Twitter as a sensor, the most related work of this paper is Sakaki et al. [5]. They have tried to detect earthquakes using a social sensor that is based on tweets. While their objective was event detection, we aimed to measure the degree of hay fever. Compared with their work, our work has a complexity in both the calculation and the evaluation of degree. This is the contribution of this paper.

We developed a system that extracts information about hay fever from tweets and that visualizes the condition of hay fever for the entire country (Japan). We provided the output of the system as the Web service, "Hay fever now!" from March 2010 to May 2010. We will introduce the system in the next section.

2 Hay Fever Observation System

The goal of the system is to generate a hay fever map similar to a weather report map. Fig. 1 is an example of output. The system first collects tweets that contain the key phrase "hay fever"[2] by using the Twitter search API[3]. Then, the system sorts tweets by location (prefecture) and visualizes them in a map.

There are some difficulties in the implementation, such as location estimation, normalization of tweet numbers and classification of tweets. We will describe them in the following sections.

2.1 Location Estimation

Every tweets are able to have geolocation information (latitude and longitude), which is very useful for the objective of our system, however, not all users give permission to provide the information. Our preliminary investigation showed that only 0.6% of tweets included geolocation information.[4] Because this is not sufficient our purposes, we used published user profile information as an alternative to geolocation information. Users can optionally register their location when registering for a Twitter account.

---

[2] Written in Japanese in the system.
[3] `http://apiwiki.twitter.com/`
[4] We investigated tweets that include the phrase "hay fever"

Fig. 1. Output of pollen observation system

Since users write the profile information in natural language, we needed to analyze the information and estimate the name of the prefecture. For the prefecture name estimation, we made a dictionary using a gazette edited by the Geospatial Information Authority of Japan[5]. The dictionary has the name of the city, ward, town, village, mountain, hill and lake for each corresponding prefecture. Most of the terms in the dictionary are written in kanji.[6] We expanded them into hiragana and the English alphabet; Twitter users sometimes wrote their location in their profiles in hiragana "さっぽろ" ("Sapporo" in English) and the English alphabet "Kanagawa, Japan". The number of terms in the dictionary was expanded from 3,196 to 11,584.

The 11,584 terms had 664 duplications. We used a Web search engine with the query "prefecture name & land name" to resolve the duplications. The prefecture name with the highest number of search results was selected as corresponding land name for each duplicate land name.

We attempted to use bi-gram similarity for the matching between the dictionary and profiles. Some users wrote their locations in the profiles as "i'm living tokyo now" which should be matched as "Tokyo".

2.2 Normalization

The popularity of Twitter is different in each prefecture. Moreover, the average number of tweets posted by one person in a prefecture may vary from the others. Fig. 2

---

[5] `http://www.gsi.go.jp/KOKUJYOHO/gazetteer_j.html`

[6] Japanese has three character sets, hiragana katakana and kanji.

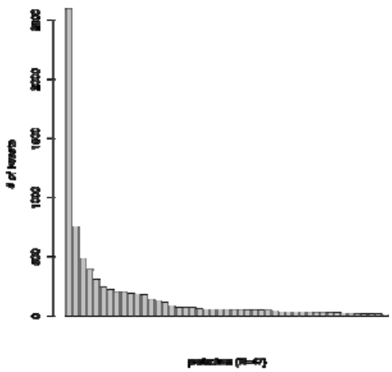


Fig. 2. Distribution of tweets over prefectures

shows the distribution of the number of tweets. Clearly we need to normalize the number. For the normalization, we collected 7,000 random tweets determined whether geolocation was added to the tweet or whether the location was estimated from the user's profile.

Using the distribution of tweets over all prefectures, we calculated a weight for each prefecture as the proportion to the max number, the number of tweets in Tokyo. The highest weight value was 259.4 for Shimane Prefecture. This meant that the number of tweets posted by users in Tokyo was 259.4 times as many as the number in Shimane Prefecture.

2.3 Classification of Tweet

If the system relied only on the key phrase “hay fever” for the data collection, it used many tweets in which the user does not suffer from hay fever. The tweets may be, “I do not want to get hay fever.” or “Am I getting hay fever? No I don’t believe that”. We attempted to classify all target tweets into two classes: getting hay fever and not getting hay fever. In order to make the classifier, we first classified 1,000 tweets by hand into five levels as described in Table 1. It was a difficult task to classify tweets into the binary classes of “hay fever” or “no hay fever” even by person because many of the tweets were vague. Five classifications were more suitable for this reason.

We used a classifier developed by Iwakura at el. [6] for the classification. The classifier was trained with 1,000 examples of (1) and (2) as positive examples and (3), (4) and (5) as negative examples. This means that the classifier organized the tweets into two classes. The accuracy obtained after five fold cross-validation was 77.27%. The classification was not an easy task. It was difficult to discriminate between suffering from hay fever or not from only the surface information of text messages. While model selection or parameter tuning may give higher performance, we used the classifier for our experiment because the improvement was estimated as minimal because of the difficulty of the task.

Table 1. Category of manual classification

Level	Number	Description
(1)	430	Hay fever with symptoms
(2)	144	Might be hay fever with symptoms
(3)	119	Hay fever but no symptoms
(4)	67	Might be hay fever and no symptoms
(5)	240	Not hay fever and no symptoms

3 Evaluation

3.1 Evaluation in Location Estimation

Precision of Location Estimation. In the experiment, the system estimated location information for 70.4% of users with the algorithm described in Section 2.1. Table 2 shows the results of the estimation for 200 random sampled examples.

The precision of the estimation is 74.5% (149/200). The examples of "Fail" occurred for complex expressions as shown in Table 3. Example expressions of "Meaningless" are "here", "town" and "in my home" from which the system should not estimate a location. In order to prevent an incorrect estimation, we can use (1) a threshold in the bi-gram matching score, and (2) the language model, which estimates the inadequacy of expression as location.

Recall of location estimation. In order to evaluate the recall of the location estimation, we checked location expressions for 29.6% users for which the system could not estimate their location. Table 4 shows the results of the evaluation for 200 random sampled examples.

The number of expressions for which the system should provide location information is 18, the sum of the number for "Detailed local" and "Slang" in Table 4. This rate is 9.8% out of the 200 samples.

Table 2. Evaluation of precision for location estimation

Category	Number
Correct	149
Fail	16
Meaningless	35

Table 3. Failure example of location estimation

Example expression	Corresponding expression in English
ポロサツシティ	"YorkNewCity" (for "New York City")
(竜)宮城	brand new youk
大阪と京都といったりきたり	Back and forth between Osaka and Kyoto

Table 4. Evaluation of recall for location estimation

Category	Example explession	Number
Detailed local	Katsushika ward, Edogawa ward	15
Slang	しぞーか, ちーば [7]	3
Wider area 1	Kanto region, Kyushu	13
Wider area 2	Japan, earth	40
Meaningless	strange adventure, “home”, “room”	129

The results of the above evaluation are as follows:

- Rate of appropriate location written in the user profile was 0.607 (0.704 * ((149+16)/200) + 0.296 * (18/200)),
- Rate of correct estimation by the system was 0.524 (0.704 * ((149)/200)),
- Recall of location estimation was 0.863 (0.524 / 0.607).

Consistency in Locations between geolocations and user profiles. The locations estimated from user profiles may be different from the locations where the user tweeted because Twitter is used often on mobile devices. In order to evaluate consistency, we compared actual locations obtained by geolocations with estimated locations from user profiles. Since 195 users posted actual geolocation, the comparison shown in Table 5 could be attempted.

In the comparison, actual locations might be multiple, and estimated location was one for one user. When (1) and (2) are regarded as correct in Table 5, the consistency is 72.8 % (142 / 195). Adding (3) as a correct, the consistency is 85.6% (167 / 195). Remembering that only 60.7% users indicated the appropriate location in their profiles, this ratio is acceptable.

Table 5. Location difference between geolocations and user profiles

Category	Number
(1) One location, and the same	132
(2) Multiple location, and includes the estimated one	10
(3) Different (neighbor)	25
(4) Different	28

3.2 Comparison with Real Sensor Data

Real Sensor Data. Using the system output generated from tweets, we discuss the comparison with actual pollen data gathered by real sensors. This analysis is based on the correlation between them. If a significant correlation exists, we can say that tweets can be used as an alternative to real sensors.

We used actual pollen data gathered by real sensors for the comparison. The pollen data were provided by a pollen observation system[8] developed by the Ministry of the Environment Japan. This system provides pollen data gathered from observatories in

[7] The corresponded expressions are; NYork for New York, boostoon for Boston

[8] http://kafun.taiki.go.jp/

all 46 prefectures, except Okinawa. Because each prefecture has 2 to 10 observatories (average: 3.1), the average value of the multiple observations for each prefecture was used for the comparison.

Correlation between Pollen and Tweets. We first calculated the value of pollen data and tweet data by day. Then, we calculated the correlation coefficient between the two data sets over the experimental period (Feb. 14, 2010 – Mar. 5, 2010) for each prefecture. Table 6, Fig. 3 and Fig. 4 show the results. This analysis was conducted for the two data sets: without classification and with classification. Both results had a weak correlation for mean, 0.145 and 0.153, respectively. The classification seems to contribute to the correlation coefficient, however, the results of a statistical hypothesis test (T-test) showed that there is no difference between these two values. This result confirmed that text classification did not contribute to a correlation between real pollen data and tweet data.

Effect of the Number of Tweets and the Amount of Pollen. There is considerable divergence in the correlation coefficient among prefectures as shown in Fig. 3 and Fig. 4. In order to analyze the effect of the number of tweets and the amount of pollen on the divergence, we confirmed the correlation between them.

Table 6. Summary of correlation coefficient for prefectures

	w/o classification	w/ classification
Min.	-0.184	-0.202
1st Q	0.00323	-0.0378
Median	0.0992	0.125
Mean	0.145	0.153
3rd Q	0.248	0.249
Max	0.646	0.703

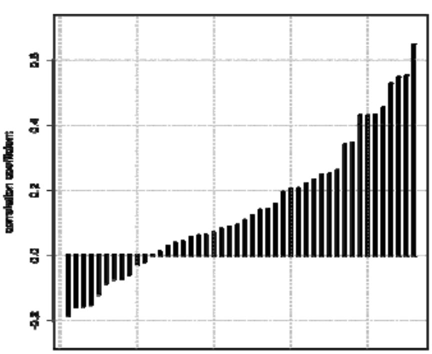


Fig. 3. Correlation coefficients for prefectures (w/o classfication)

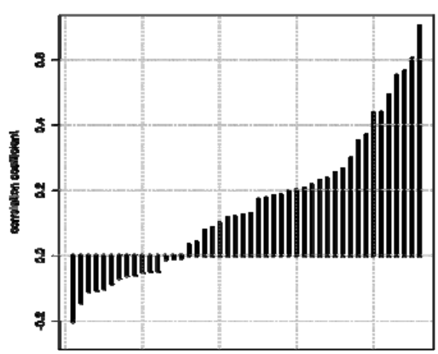


Fig. 4. Correlation coefficients for prefectures (w/ classfication)

The result of Pearson's product-moment correlation test for the number of tweets was a 0.31 correlation coefficient with a 0.05 level of statistical significance for data without classification, and 0.33 for data with classification. This result confirmed that the more tweets that are posted, the higher the correlation coefficient between real pollen data and tweet data. In other words, the more tweets that are posted, the higher the possibility that Twitter can be used as an alternative for real sensors.

We also attempted the same analysis for the amount of pollen, however, there was no significant correlation between them.

Location Distribution of Correlation Coefficient. Fig. 5 shows the distribution of the correlation coefficient over a map of Japan for each prefecture. Dark color means high value of the correlation coefficient. While prefectures that have high correlation coefficients were centered around the Kanto and Chugoku regions, prefectures that have low correlation coefficients were centered around the Kansai region. The mean value of correlation coefficients in the Kanto region is 0.434, which is significantly higher than the mean value of all prefectures (0.153). On the other hand, the mean value of the correlation coefficients in the Kansai region is -0.0108 that is a significantly small value.

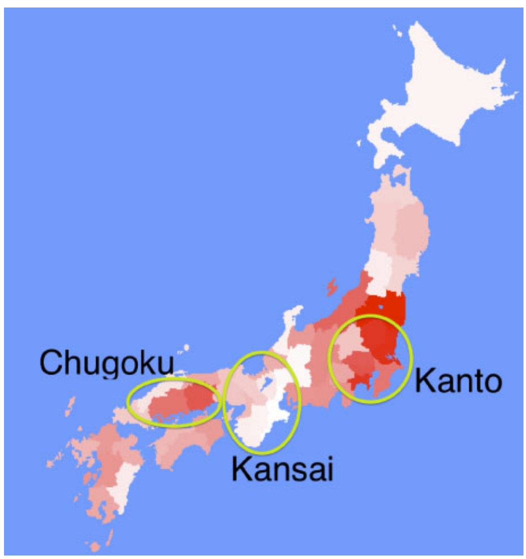


Fig. 5. Distribution of correlation coefficients on a map

4 Discussion

4.1 Method of Evaluation

Evaluating the output of our system was very difficult. We estimated the degree of hay fever in this experiment. This estimation was based on the hypothesis that the more people who showed symptoms of hay fever, the more tweets that would be posted. In order to evaluate the estimation properly, it should be compared with data on the symptoms of hay fever instead of on pollen number. However, it was difficult to obtain data, such as the number of patients with hay fever. We used the pollen counts as an alternative indicator of symptoms of hay fever and investigated the correlation between the count and the degree, which was estimated from tweets. We found a correlation between them in several areas, however, the lack of a correlation in the rest of the areas might be due to the problem of the aforementioned hypothesis. Even if the pollen count were the same in two areas, the symptoms of hay fever might be different because of other conditions, such as temperature, humidity and wind. The evaluation should be refined by involving the different kinds of data.

4.2 Degree of Symptoms in One Tweet

In this experiment, all tweets were calculated equally, except document classification, which we mentioned in Section 2.3. This meant that the degree of symptoms in each tweet was not determined, however, actual tweets had different degrees as shown in the following tweet examples:

- My eyes are itchy from hay fever,
- Sneeze, mucous, itchy eyes, this is surely hay fever,
- My eyes are so irritated! Hay fever!

Since we tried to measure the degree of phenomena, such as hay fever instead of event detection as mentioned in Section 1, we should extract information about the degree from each tweet. Extending the classification described in Section 2.3 from the binary classification to a multi-class classification can carry out this objective, however, we need a more accurate classification method than the current one (77.27% accuracy) because the high number of classes causes a reduction in accuracy.

5 Conclusion

This paper verified that Twitter could be used as an alternative to real-world sensors. The results showed that there was a positive correlation between pollen data and tweet data, and Twitter has the possibility of being used as a real-world sensor at least in a particular area such as hay fever. Our experiments proved that the number of tweets affected the correlation. This result confirmed quantitatively an intuitive interpretation that the number of tweets leads to high performance as a sensor. Our experiments also showed the limit of our algorithm, such as the divergence of the correlation coefficient among regions. We must seek the reason of the divergence and must involve them into the system architecture in order to make the sensor accurate.

References

1. Pang, B., Lee, L.: Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval A1 2(1-2), 1–135 (2008)
2. Java, A., Song, X., Finin, T., Tseng, B.: Why we twitter: understanding microblogging usage and communities. In: Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web mining and Social Network Analysis, WebKDD/SNA-KDD 2007, pp. 56–65. ACM, New York (2007)

3. Kwak, H., Lee, C., Park, H., Moon, S.: What is twitter, a social network or a news media? In: Proceedings of the 19th international conference on World Wide Web, WWW 2010, pp. 591–600. ACM, New York (2010)
4. Naaman, M., Boase, J., Lai, C.H.: Is it really about me?: message content in social awareness streams. In: Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work, CSCW 2010, pp. 189–192. ACM, New York (2010)
5. Sakaki, T., Okazaki, M., Matsuo, Y.: Earthquake shakes twitter users: real-time event detection by social sensors. In: WWW 2010: Proceedings of the 19th International Conference on World Wide Web, pp. 851–860. ACM, New York (2010)
6. Iwakura, T., Okamoto, S.: A fast boosting-based learner for feature-rich tagging and chunking. In: Proceedings of the Twelfth Conference on Computational Natural Language Learning, CoNLL 2008, pp. 17–24. Association for Computational Linguistics, Morristown (2008)

Reacting with Care: The Hybrid Interaction Types in a Sensible Space

Guo-Jhen Yu[1] and Teng-Wen Chang[2]

Graduate School of Computational Design,
National Yunlin University of Science and Technology
{g9734703,tengwen}@yuntech.edu.tw

Abstract. Our family members are used to interact with virtual characters than real world. One way to solve this problem is to stop using computers. The other solution is a usable communication platform that can take advantages of both worlds. With the spaces equipped with the media technology defined above, what are the enhanced interfaces for supporting family members communication? We developed the information transmission framework for the refined co-existing spaces that we called hybrid interactive co-existing spaces (HyCoe spaces). The result for implementation is tremendous due to the complex of sensible spaces and the implicit intention during the communication among family members.

Keywords: Sensible Environment, Interactive Co-existing Spaces.

1 Introduction

Computing has become an important sector of modern daily life. The communication of family is gradually depended on the computing platform. Our family members are used to interact with virtual characters than real world. One way to solve this problem is to stop using computers. The other solution is a usable communication platform that can take advantages of both worlds. Since we cannot afford not using computing apparently, the proposed platform is to find a suitable computing solution to enhance the issues addressed.

1.1 The New Interactive Behavior of Family

As mentioned above, the computing used in daily routines has influenced the communication behaviors in the family context. With observation over everyday's event, the behaviors by the family members such like: walking, turning up and off consumer electronic and so on, although are belonged to individual actions, but they may still affect other members' and often hard to distinguish which one is more effective than others. If we need to provide a space for enhancing communication among family members, we need to know which one is suitable to be accomplished with media.

Home is filled with memories and familiar things that family members are living in it and sharing everything with others[1]. In family interaction, information exchanging and transferring are the important facts for gluing family together. Simply

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 250–258, 2011.

speaking, family's common memory is based on those exchanged information. The studies[2] have divided members' activities into individual and social activities; each one can be further classified as media interaction and spatial interaction. Social activity is family members interacting with other members through media or spatial interaction, such as leaving reminding text messages. In between individual and social activities, user needs media and space to get message transfer and feedback, so in member various interaction mode, we need various interfaces to accommodate these diverse situations.

1.2 Sensible Environment

In the computing concept, especially the ubiquitous computing, often used for developing social application for spatial interaction, space containing countless sensors, all static objects and space itself are all part of a seamless Human Computer interface between habitants and their surroundings. This interface or called sensible environment can sense habitants' condition, behaviors, environmental fact, and the contextual information[3, 4]. The individual activities as defined before objects can sense the requests from human directly and provide direct feedbacks such as turning up the light switches, the light goes up. Nevertheless, individual activities are not that direct in social activities.

Context aware researches are into this kind of problems. Dey defined the context as a situation that can be used to describe characteristics of information, and this information is be considered of users and their application in the mutual interaction [5]. Further, computing will record habitants' behaviors according to the timeline, sets previous tasks as the target, and analyze the activities in the spaces.

2 The Problem

With the spaces equipped with the media technology defined above, what are the enhanced interfaces for supporting family members communication? Commonly speaking, the interface might lie upon the analysis of the interactive behaviors in family daily routines. We then further narrow down the problem into developing a space that can utilize the interactive behaviors of daily routines. This space is ubiquitous and sensible for understanding the behaviors and reasons behind them.

Since the exploration nature of this problem addressed above, we reviews related literatures for the specific interaction behaviors needed in our context as well as how to develop a prototype to testify the concept gained from the reviews. Finally, the evaluation is done in implementation and discussion will carry on for uncovering the lessons learned in the process.

3 Literature Reviews

Following the concept of ubiquitous computing in spaces and family interactive behaviors with media, we are reviewing the possible interaction modes appeared in the environment and how to integrate them into the functional analysis for the interaction design prototype of this research. First we need to examine the concept behind the coexisting spaces and the interaction modes in them.

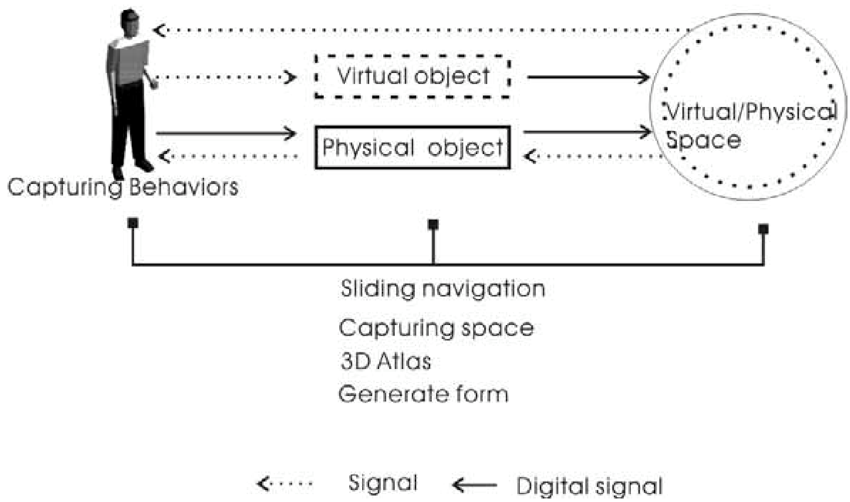


Fig. 1. The four interactive activities between User and Virtual/Physical Spaces in co-existing spaces [7]

3.1 Co-existing Spaces

In the sensible environment, presentation and collection of data can be divided into physical and virtual means of interaction patterns that help computers better understand the needs of the users also can give users a more appropriate feedback. In 2004, Lai and Chang proposed a coexistence of the real and imaginary information transfer for the structure of information processing in the classification of the data after, and then through the digital media present in the physical environment and virtual environment[6].

Lu in 2005 [7] described a co-existing space containing the virtual and physical media and the state space as shown in Fig. 1. The interactive activities in co-existing spaces is divided into four: (1) physical interaction between media of physical space, (2) entities of the virtual space, interactive media, (3) The impact of virtual media physical space of interaction, (4) virtual media of virtual space of interaction. Virtual space of freedom of information delivery, and physical manipulation of the intuitive input and output, combine to form an integrated virtual and physical state.

3.2 Three Sensing Types: Direct, Indirect, Slow

While classifying the interaction preparing for the prototype, three sensing types are discovered with the interaction literatures: direct, indirect and slow. Direct and indirect sensing types refer to how data or commands are sensed into a system[8].

(1) Direct sensing: System will display the implicit information visually to the users, and users will direct manipulate the information available on the interface [9]. By understanding the need of the people, direct sensing requires objects that are controllable by people for the immediate feedbacks, for example: people control the mouse, the system gives back the feedbacks in computer screen. Direct devices have no intermediary; the movement of the body equals the input to the machine[10].

(2) Indirect sensing: indirect sensing device converts/analyzes sensed human behaviors data into reasons. Such as the environmental sensors will gather people's needs, and may reflect back on the environment. For example: People are aware of the cold environment, the system adjusted to control air temperature. This interaction would not disturb people [10].

(3) Slow sensing: Slow sensing is based on giving feedbacks long after the event occurred that is often used in a more reflective environment. Not just delay feedbacks but changes that will take time for users to reflect and to realize the consequence of an activity. Such technology is slow in nature[11].

These three sensing modes: direct, indirect and slow sensing, along with the fourth interactive mode that Lu has proposed above form the basic interactive modes in this paper.

3.3 Three Display Types: Direct, Indirect, Slow

With three sensing types, the output of an interaction is also examined. We will analyze the cases below to explain three display types in our scope: direct, ambient and slow displays.

(1) Direct Display represents most of visualization of interface nowadays that incorporate both virtual and physical information. For example, the virtual recipe done by Selker in 2005 [12] represented the virtual digital information on the cabinet in front of the users besides the physical kitchenware. While using the symbolic visualization to remind users to follow the direction and to operate the recipe steps, such information transfer mode with the contents of specific information presented (shown in Fig. 2). Transformation process in the interactive information in a way directly to the user, the user can manipulate the information directly. The digital information is not only in the display panel but also displayed in the surroundings.

Fig. 2. Virtual information displays in the physical environment

(2) Ambient Display on the other hand represents subtle visualization for indirect interaction. The famous case AmbientRoom [13] took living space as a sensible space with the devices embedded in the wall, ceiling or desktop. It provides users with the information via light or sound (as shown in Fig. 3). Ambient Display is a new way of

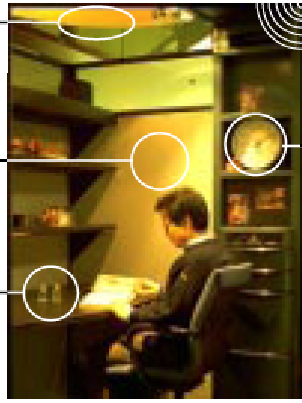


Fig. 3 Ambient displays and controls

display, this display no longer requires users' attention, but embedded in the environment awaiting for the attention of users. Environment is often used as a medium for transmission of information. The purpose of this type of display is not disturbing people's attention and allows users to focus on their current tasks while still bring the information cross.

(3) Slow Display: Slow Display is a media that will take a longer time to notice the feedbacks. The usual media in Slow Display is nature materials such as plant or grass. Within these media, PlantDisplay takes a step forward by combining the concept of ambient display that the idea of changing the interface to the plant shows feelings and emotions of people. Users at plant interface, the interface will be sent out by the timing of the meaning behind their move and cause a deep impression. Transformation process in the interactive information, the system of virtual digital signal into meaningful information, the system entity to control the plant given the amount of light and [14]through a long time Accumulation of information[15, 16].

3.4 Summary

With the three cases in 3.3 to the interface type, media, and the presentation of information is shown in Table 1. We can find that information is presented the same sense as the information can be divided into direct, indirect and slow presentation.

Table 1. The interface types/media/information of three Cases

Case	Interface type	Media	Information
Smart kitchen	Virtual	Environment	Direct
AmbientRoom	Physical Virtual	Environment	Indirect
PlantDisplay	Physical Virtual	Object	Slow

4 Implementation

Based on the concept above, we developed the information transmission framework for the refined co-existing spaces that we called *hybrid interactive co-existing spaces* (HyCoe spaces). The basic framework and its components is described as below, and followed by the prototype of HyCoe.

4.1 Framework

With the analysis above, the framework of HyCoe spaces is illustrated in Fig. 4. Based on the co-existing spaces defined in [6], this framework contains three folds: two information processes interacting with two spaces, three interaction modes, and user activities. While the process triggered by users' activities in space, system will employ different mode of interaction with corresponded contextual interplay.

The information transmission of HyCoe spaces is divided into direct, indirect and slow. Four interplays based on information transmission are: collecting related information, interpreting collected information, integrating interpreted information and communicating with others and back to the users.

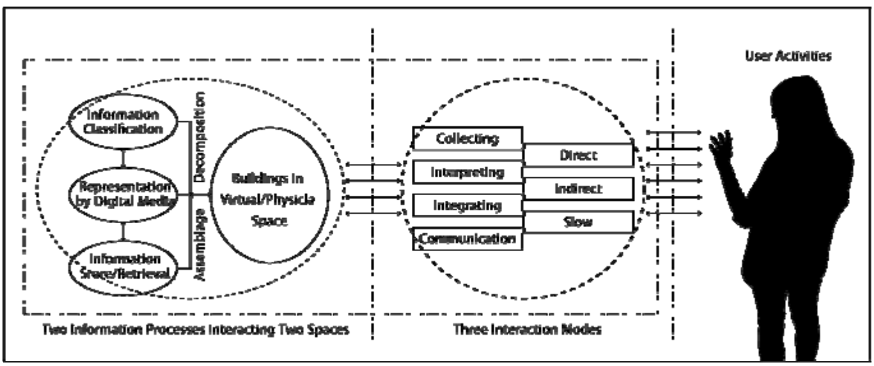


Fig. 4. The framework of HyCoe Spaces

4.2 Hybrid Interactive Co-existing Spaces

The prototype of HyCoe Spaces is built on the physical environment of SOFTLab in NYUST. The space is equipped with sensors that can collect the activities data directly from the habitants and connecting to the server. The sensors used are distance measuring sensor, sound sensor, flex sensor, RFID tags and reader (as shown in Fig. 5). Distance sensor can detect the distance between habitant and the coffee table that can notice who is coming close to the coffee table. Flex sensor can detect when someone is sitting in the sofa that can also cooperate with distance sensor for detecting a more complex situation such as whether anyone is coming close the coffee table and sitting in the sofa. The RFID tags are attached to the daily objects such as tea-cups, plats and books. When habitants put an object with RFID tag on the table, the reader can detect which one is on the table; further server can use this information to build up a reason behind these actions such as when table just detects a book is in might mean habitant would like to read a book.

(a) (b)

Fig. 5. (a) The activities of habitants, (b) flex sensor and microcontroller (arduino)

(a) (b) (c)

Fig. 6. (a) projected screen, (b) ambient display, (c) plant display

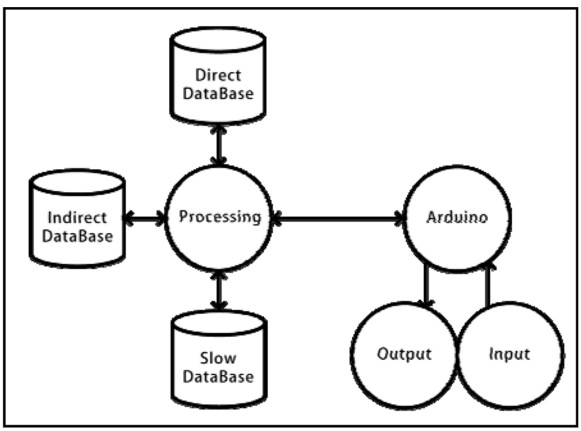


Fig. 7. System Architecture

After setting up the sensing environment, the display will then be implemented including virtual projected screen, ambient lights and plant display (as shown in Fig. 6). Virtual projected screen gives habitants the direct feedback of information that will take the input from habitants as well. The ambient lights show the current condition of habitants. For example, the light went red when habitant is working. Plant display on the wall reflects the friendliness of habitants nearby.

For the software platform, we develop the prototype on two packages: processing (http://www.processing.org) and arduino (http://www.arduino.cc) that are open sourced computing platform that is often used in interaction design. Arduino collects the digital/analog information from the sensors and send them to the server. While received this information, Server applies visualization or store the data into database (MySQL, an open sourced database management system) via processing. The simplified structure of HyCoe spaces prototype is shown in Fig. 7.

5 Discussions

In the implementation of HyCoe Spaces, we apply variety of sensors onto different furniture. Due to the explorative nature of prototyping, we constantly need to increase or alter different sensors with different combination of furniture such as coffee table used in the living room. Consequently, we design the sensors with module-in mind as much as we can. The sensor module in the coffee table is shown in fig. 8. In fig. 8, it shows the modular concept of sensors that are attached with the concept of latches that can allow the sensors modules be easily altered and combined under different furniture. The tracks system of modular design also let us to change the position of sensors with ease.

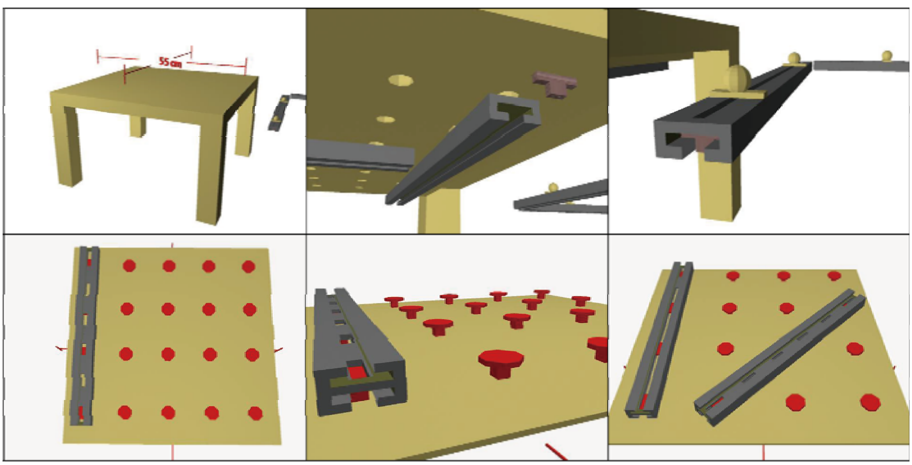


Fig. 8. The design of sensor module of coffee table

6 Conclusion

In this research, we studied an enhanced interface for supporting family members' communication via analysis of interactive behaviors of daily routines. With the analysis, a framework of hybrid interactive co-existing spaces is proposed and a prototype is implemented for testifying the computational framework. Four interplays with three interactive modes: direct/indirect/slow are uncovered and frame the possible interaction. The result for implementation is tremendous due to the complex of sensible spaces and the implicit intention during the communication among family members. With this platform developed, a further experiment of family scenario and communication can then be carried further in the next step.

References

1. Silverstone, R.: Why Study the Media. Sage Pubns, Thousand Oaks (1999)
2. Chen, C.-W., Chang, T.-W.: The study of Context-Oriented Family Inter-Relationship Platform. In: NEW FRONTIERS - 15th CAADRIA Conference 2010, Hong Kong (2010)
3. Bannon, L.J.: A human-centred perspective on interaction design. Springer, London (2005)
4. Beigl, M., Gellersen, H.-W., Schmidt, A.: MediaCups: Experience with Design and Use of Computer-Augmented Everyday Artefacts (2001)
5. Dey, A.K., Abowd, G.D., Salber, D.: A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications. Human-Computer Interaction 16 (2001)
6. Lai, I.-C., Chang, T.-W.: Dynamic Interactions Between Users and Information in a Co-existence Space. In: Orbak, i.H. (ed.) The eCAADe 2004 22nd Conference, Copenhagen, Denmark, pp. 58–65 (2004)
7. Lu, K.-T.: CapX: Co-existing Design Capturing Device. Graduate Institute of Architecture College of Humanities and Social Science. Master in Architecture. National Chiao Tung University, Hsinchu (2005)
8. Jacob, R.J.K.: Human-Computer Interaction: Input Devices. Acm Computing Surveys 28 (1996)
9. Shneiderman, B.: Direct manipulation for comprehensible, predictable and controllable user interfaces. In: 2nd International Conference on Intelligent User Interfaces, New York, NY, USA (1997)
10. Mclaughlin, A.C., Rogers, W.A., Fisk, A.D.: Using Direct and Indirect Input Devices: Attention Demands and Age-Related Differences. ACM Transactions on Computer-Human Interaction (TOCHI) 16 (2009)
11. Hallnäs, L., Redström, J.: Slow Technology – Designing for Reflection. Personal and Ubiquitous Computing 5 (2001)
12. Selker, T.: Smart Kitchen: The Kitchen of the future. ARCHIDATA, Taiwan (2005)
13. Wisneski, C., Ishii, H., Dahley, A., Gorbet, M., Brave, S., Ullmer, B., Yarin, P.: Ambient Displays: Turning Architectural Space into an Interface between People and Digital Information. In: Yuan, F., Konomi, S., Burkhardt, H.-J. (eds.) CoBuild 1998. LNCS, vol. 1370, p. 22. Springer, Heidelberg (1998)
14. Kuribayashi, S., Wakita, A.: PlantDisplay: Turning Houseplants into Ambient Display. In: ACE 2006, Hollywood, California, USA (2006)
15. Easterly, D., Kenyon, M.: Bio-Fi: Inverse Biotelemetry Projects. In: MM 2004, New York, New York, USA (2004)
16. Shih, J.-H., Chang, T.-W., Hong, H.-M., Li, T.-C.: Physical representation social presence with interactive grass. In: International Conference on Human-Computer Interaction (2007)

GoCoBa: Interactive Installation Design Applied on Combination of Context and People

Jia-Xuan Zhan and Kuo-Kuang Fan

Graduate School of Computational Design, National Yunlin University of Science and Technology, 123 University Road, Section 3,
Douliou, Yunlin 64002, Taiwan, R.O.C.
{g9834709,fankk}@yuntech.edu.tw

Abstract. The combination of people and context is easy to be neglected in the Human-Computer Interaction (HCI). Context always affects unconsciously behaviors of people, and it is inessential for people. Though observing realistic environment and analysis, we aim to design a system for context, object, and space interaction in sports bar. We also search some interactive installation cases to find the related cases in order to understand the possible interactive patterns. According to case and behavior practice to design prototype, we made GoCoBa, an interactive design for bar goers to immerse in the context. To sum up, GoCoBa system using cup as physical computing and context computing help bar goers have interaction with the context.

Keywords: Interaction Design, HCI, Sports Bar.

1 Introduction

The combination of people and context is easy to be neglected in the Human-Computer Interaction (HCI). Context always affects people unconsciously behaviors, and it is inessential for people. Because context is unnoticeable, designers may not pay much attention to it while designing. Though HCI technology, designers can help people and context have combination, and build intelligent space which could display information, transfer data, and participate environment development. For people and object in the context, sensor of interaction design can sense and give feedback and help[1]. How can HCI help to integrate people and context is the main idea of this research.

1.1 People in Sport Bar

In this problem, we need to find a suitable context to explore the possible interaction. Sports bar is not only for sport enthusiasts to meet and enjoy the games, but also provides a space for people to relax, have drink, and watch sport games. Comparing with watching games at home; staying in the bar is more interesting because of the environment and atmosphere. In order to solve the problems between people and context, we observed the realistic environment to find the problems and possible solutions. We decided to adopt sports bar as the context of this research.

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 259–267, 2011.

1.2 Physical Computing

For solving the interaction in a context that combines both people and surroundings, we adopt a computing technology called physical computing. In the context often search a real object in the sports bar as the computing object to collect physical interaction data. This object could be anything such as a chair, desk, floor, wall, or cup, and it will play an important role to identify relationship between people and context. In order to achieve this interaction design, we found cup is an ordinary but significant object in sports bar, and cup is often applied as a physical computing device, in many interaction design cases.

2 Finding the Problem

Whole the object is decided, the interaction between this object and people are identified as observation. Physical site we investigated with the focus on the interaction between people and their cups. Several instances appeared during the observation sessions, and the engagements of events are recorded for analysis.

There are (a) Person A entered the bar, ordered a cup of drink, watched ball game alone, and sipped of the drink when the game was time out. (b) Person B and a friend went to the bar. Both of them ordered a drink individually. They watched the game together. They clinked the cups to cheer, whenever there was a beautiful play or score. (c) Person C entered the bar and ordered a cup of drink as well. Walking into people to have chat, Person C also clinked the cup when having good play in the game. In the observation, we found that people did not have much interaction, and the context did not emerge its specialty. Therefore, interaction technology might be a method to help people and context to have good connection, and it can also help bar goers have more chances to communicate with friends or strangers.

3 Case Studies

Though observing realistic environment and analysis, we aim to design a system for context, object, and space interaction in sports bar. We also search some interactive installation cases to find the related design cases in order to understand the possible interactive patterns.

3.1 iBar

iBar (2007) by the Landon based company Mindstorm Ltd and its partners, is using a bar table as an interface. When some objects put on iBar, there will be some flame shown around the single object. Besides; objects and objects will be light connected. In the case, the light can catch people's attention, make people connected, and arouse bar goers' imagination and creativity. While drinking and chatting, people can also have abundant interaction[2], as shown in Fig. 1.

(a) (b)

Fig. 1. iBar (a) Cups on the table having connection. (b) Bard goers having interaction though iBar.

3.2 Lover's Cup

Chung made an emotional interactive installation device called Lover's cup in 2006. Two cups used in two different places for lovers to receive the information of the other person though the action of using one of the cups. The sensor is installed in the device, once a lip touches one of the cups, the sensor will be active and sent information to the other cup. By the time the other cup receives the data, the cup will shin and the show the scale so that the user can easily understand the other one is currently drinking. This device Integrates into daily activity incidentally[3], as shown in Fig. 2.

(a) (b)

Fig. 2. Lover's Cup (a) Though interaction of Love's cup, when one is drinking, the other cup will display the scale. (b) Diagram of Love's cup device.

3.3 Ada-Intelligent Space

Ada-intelligent space is a case of interactive floor designed by Kynan Eng in 2003, using the pressure sensors on the floor. Whenever people step on, the hexagon-shape device will light up as feedback. This system covers four basic behavior interactions. The light will follow every users and focus of one of the users. Besides, the device gives users cuing, when the conditions are reached, the game will be start[4], as shown in Fig. 3.

(a) (b)

Fig. 3. Ada-Intelligent Space (a) the left-top is the picture of computing; the others are bar goers in the context shot by camera. (b) The lighting is feedback from the floor.

3.4 Analysis

In the case of analysis, we found those designs have their disadvantages. iBar only benefits to people around bar, cannot cover the whole environment. Lover's cup is mainly for lovers; its intimacy interaction pattern is not suitable for everyone in the bar. Ada's interactive floor can be used for whole space, but it does not have much interaction system for users to connect each others. The cases analysis as followed the Table 1.

Table 1. Case analysis

Case	Context	Item	Behavior	Feedback
iBar	Bar table	Cup, hand, phone, cards, etc.	Put objects on the table	Table lights and focus the objects
Love's Cup	None	A couple of cups	Drinking	Light spots on the cup to show the scale
Ada- Intelligent Space	Space	Floor	Step on the floor	Lights on the floor and reflex the position of people

By case analysis, we concluded the advantages of those three cases. We would like to have people join the interaction as Ada's interactive floor did. We also took the concept of iBar, using light to direct people to have interaction. Moreover, we use cup as controller, like lover's cup, using cup as an input interface to connect people's behaviors.

4 The System: GoCoBa

According to case and behavior practice to design prototype, we made GoCoBa, an interactive design for bar goers to immerse in the context. The action of drinking enlightens our design interactive installation device. Bar goers can have interactive

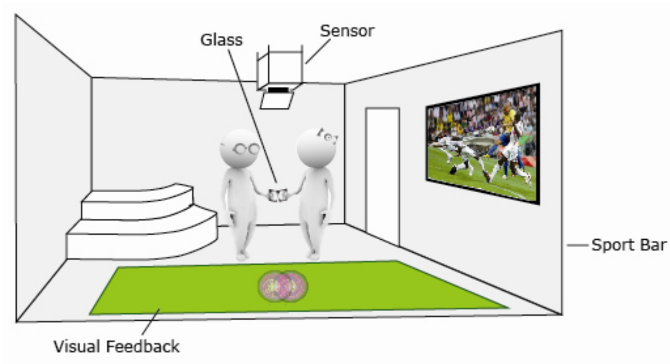


Fig. 4. GoCoBa interface diagram

communication in anyplace in the sports bar, and sending response feedback to the context. The Fig. 4 The diagram of GoCoBa interface to show interaction between people and context.

4.1 System Design

The system of GoCoBa is comprised of:(1) Input: receiving message by utilizing IR sensor, and transfer the message to computer.(2) Computing: the computer receive the message of light spots, and transfer them into location data, and finally categorizing in order the data into different database.(3) Output: projecting the processed message on the context as visual feedback.

When bar goers enter sports bar and take the cups, (input) the input sensor (IR sensor) will receive infrared rays from bottom of cup. (computing) This time information will be sent to computing unit to compute. To start with, the system takes information from bar goers and transfer into location data. And then, the location data will be saved in data base for loading. Finally, information from database sends to rule base to compare, and the result will export. (output).

Though generalizing from database, there are four outputs: Act 1, Act 2, Act 3, and Act 4. These four outputs correspond to situations of users, displaying different visual feedbacks, as shown in Fig. 5.

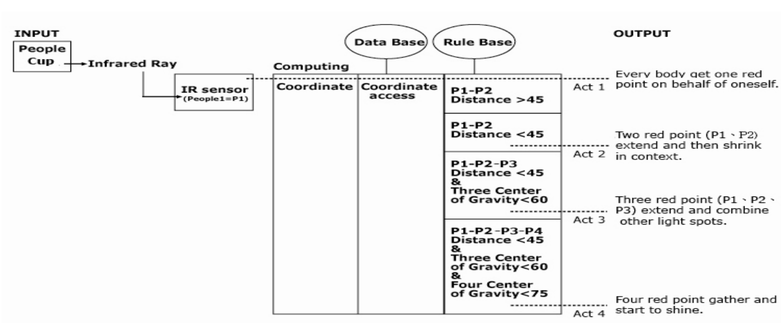


Fig. 5. System schematic diagram

4.2 Components

IR Sensor induces the IR signal from cups, and the infrared ray LED on the bottom of cup is the input of HCI. Infrared rays LED is not easy to be affected by environmental lights. Besides, infrared ray is small and light, so it is easy to be installed on the bottom of the cups. 12V LED as cuing light and battery are installed on the bottom of the cups. Bar goers can have informal interaction by the device, as shown in Fig. 6.

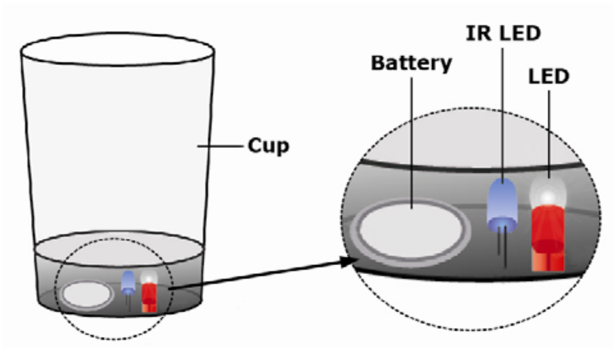


Fig. 6. Install IR LED light and 12V LED lights in the glass

4.3 The Interactions

If the distance of cups is different, the feedback will be different. We utilize projection on floor to provide feedback. According to analysis result, there are four kinds of responses as followed: (a) When there is only one point, the system will

Table 2. user behavior on the system input and output

People	Input	output	Condition
(a)One	P1	L1	Point
(b)Two	P1, P2 Distance => 45cm	L1, L2 (P1+P2)/2	
(c)Three	P1, P2, P3 Distance => 45cm Central =>60	L1, L2, L3 (P1+P2+P3)/3	
(d)Four	P1, P2, P3, P4 Distance => 45cm Central => 75cm	L1, L2, L3, L4 (P1+P2+P3+P4)/4	

P1, P2, P3, P4 = people coordinate
L1, L2, L3, L4 = light coordinate

measure the distance. If the point is 45cm far way from other point, the point will have its own light spot. (b) If there are two points and their distance is less than 45cm, the light spots will extend and then shrink. (c) If there are three points and the distance between three points is short than 45cm and the distance is less than 60cm, the light spots will extend and combine other light spots. (d) When there are four points and the distance of four points is shorter than 45cm and the distance from center to each point is less than 75cm, the light spots will gather and start to shine, as shown in Table 2.

5 User Testing

Users' background description. (Table 3.)

Table 3. users' background

User id	Age	Gender	Bar experience	profession
User A	23	male	Yes	Student
User B	24	female	No	Student
User C	24	Male	No	Student
User D	34	Male	Yes	Student
User E	42	Male	Yes	Teacher
User F	29	Male	Yes	Student

Testing progress

(a) Bar goers enter the context.
(b) Bar goers take interactive installation cup.
(c) Bar goers had interactive behavior.
(d) Take interactive result as feedback to the context.
Refer to Fig.7.

Fig. 7. GoCoBa interface prototype. People in context. Using interactive installation design. Interface of interactive installation design.

Testing result

By the time we had 6 users tasted, we concluded the result. We found that because of the age and life experience difference, the responses of interaction are slightly unlike. Therefore, we also had interviews to have deeper understanding of users' feeling, as shown in Table 4.

Table 4. users' testing result

	User A	User B	User C	User D	User E	User F
Progress (a)	Standing and watching game with F	Currently leaving this context	Sitting on the sofa and watching game with E	Watching game alone	Sitting on the sofa and watching game with C	Standing and watching game with A
Progress (b)	Holding a cup and finding the small red point following	entering the context and finding the small red point following	Holding a cup but did not notice the small red point	Holding a cup and finding the small red point following	Holding a cup but did not notice the small red point	Holding a cup and finding the small red point following
Progress (c)	Finding the small red point and having interaction	The small red point still following	The point of C interacting with the points of D and E	The point of D interacting with the points of C and E	The point of E interacting with the points of C and D	Finding the small red point and having interaction
Progress (d)	Chatting with F	Join the group C, D, and E and chatting	having interaction and chatting with the others	having interaction and chatting with the others	having interaction and chatting with the others	Chatting with A

Interviews and commons

- User A, F: While using GoCoBa system, the light on the floor chased me. I felt quite interesting. Besides, when I was chatting with F, mine light spot had interaction with F's light spot. That was impressive.
- User B, D: When people gathered, light spots increased with the number of people. That made me wants to join the group of people to see the changes of lighting.
- User E, C: When D and E joined us, we became a big group. This time the lights seemed like telling people to join us.
- User C: Because of the cup movement, light spots moved around that were fascinating. If the lights can represent different personalities of different people, that will be more fun!
- User D: as far, GoCoBa is only projecting on floor, I wonder if it is possible to be shown on walls?
- User F: the sensors should not only be on cups, more cameras may add into the context to sense, and to diversify the effects.

6 What We Learned

Though prototype design and observation of its work, we found that:

(a) bar goers using GoCoBa system in sports bar can immerse in the context and be willing to have more interaction with some other people.
(b) People in general think interaction feedback of light spots is interesting, and some people even feel light spots like the eager emotional to have communication with other people. HCI help bar goers in the context have intimate atmosphere.
(c) Because of the limited time, the prototype is still no prefect, and may have some problems need fixing. So far the context research cases are lack, and the ability of stimulation is deficient. Those should be considered into the following research.

7 Conclusion

To sum up, GoCoBa system using cup as physical computing and context computing help bar goers have interaction with the context. Thought the computing, cups can play an important role of HCI in the bar to make people and context integrated together. The behavior of holding cup can be computed and give feedback to context to induce the action of holding cup have integration with context.

Context and people have interactive by using HCI is worth for researchers to pay attention. As we can see, HCI help people to immerse in the context; moreover, it is much easier to give people feedback from context. HCI makes more interaction actions happen between people and people, people and context.

References

1. Guallart, V. (ed.) C.L.: Media House Project: the House is the Computer, the Structure is the Network. Architectural Design 75 (2004)
2. http://www.i-bar.ch/en/info/
3. Chung, H., Lee, C.-H.J., Selker, T.: Lover's cups: drinking interfaces as new communication channels. In: CHI 2006 Extended Abstracts on Human Factors in Computing Systems, New York, NY, USA (2006)
4. Eng, K., Babler, A., Bernardet, U., Blanchard, M., Costa, M., Delbriick, T., Douglas, R.J., Hepp, K., Klein, D., Manzollil, J., Mintz, M., Roth, F., Rutishauser, U., Wassermann, K., Whatley, A.M., Wittmann, A., Wyss, R., Verschure, P.F.M.J.: Ada - Intelligent Space: An artificial creature for the Swiss Expo.02. In: IEEE International Conference on Robotic & Automation, Taipei, Taiwan (2003)

Part III
Orientation and Navigation

Behavioral Cost-Based Recommendation Model for Wanderers in Town

Kenro Aihara, Hitoshi Koshiba, and Hideaki Takeda

National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan
{kenro.aihara,hkoshiba,takeda}@nii.ac.jp

Abstract. This paper proposes a new model for recommendation based on the behavioral cost of recommendees in town. The model is based on cost-benefit analysis of the information provided to the user, referring to the model of temporal discounting and preference reversal. Here we assume that behavioral cost may be regarded as time in temporal discounting. A recommender system based on this model can select information, which is located in the surrounding area (not so far away) and may be preferred by the user, if the system can estimate where the reversal phenomenon may occur. The experiments were made using an experimental social service, called “pin@clip”, which is an iPhone-based social bookmarking service in Shibuya, Tokyo, Japan that has been operating since December 2009. The experimental results show that the phenomenon of preference reversals might occur, even though the authors could not obtain statistically significant data.

Keywords: context-aware computing, location-based service, recommender system, behavioral cost, user modeling.

1 Introduction

In recent years, the amount of geotagged information, which contains geographical data, has been rapidly growing. GPS-equipped smartphones facilitate users to embed location data in their content, such as tweets and photos, and to post them to social services like Facebook[1] and Twitter[2]. Meanwhile, map-based online services like Google Maps[3] are getting more and more popular, and one can use them for navigation while one is roving in town.

A lot of recommender systems have been proposed for wanderers [1][2]. Almost all of them seem to be location-aware and assume that the nearer the provided information is located, the more useful it is for users. This assumption implies that the information that may be preferable for the user but located a little further away vanishes from the user because of a massive amount of information, such as micro blogs like Twitter tweets. On the other hand, although a recommender system can filter out the information that is probably uninteresting to the user based on the collaborative

[1] http://www.facebook.com/
[2] http://twitter.com/
[3] http://maps.google.com/

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 271–279, 2011.

filtering model that is broadly used in existing recommender systems on the Internet, the authors suppose that the information that may be preferable to the user must depend on the user's situation and the collaborative filtering model is not enough, especially when the user is wandering around in town.

First, this paper shows the proposed model with related models in psychology and behavioral economics. Next, the experimental mobile service for social bookmarking, pin@clip, is introduced briefly. Then, an experiment and its results are described. Finally, the validity of the model and future issues are discussed.

2 Background

2.1 Location-Based Information Services

A lot of network services with location data are proposed, and some of them, such as foursquare[4], are getting popular. Usually location information is given as geographical coordinates, that is, latitude and longitude, a location identifier such as ID for facilities in geographical information services (GIS), or a postal address. Google has launched Google Places[5], which gathers place information from active participating networkers and delivers such information through Google's web site and API (application programmable interface). Google may try to grasp facts and information on activities in the real world where it has not enough information yet even though it seems to have become the omniscient giant in the cyber world. Google already captures some real world phenomena in its own materials. For example, it gathers landscape images with its own fleet of specially adapted cars for the Google Street View service. However, the cost of capturing and digitizing facts and activities in the real world is generally very expensive if you try to obtain more than capturing photo images with geographical information. Although Google Places may be one of the reasonable solutions to gathering information in the real world, it's not guaranteed that it can grow into an effective and reliable source reflecting the real world.

Existing social information services, such as Facebook and Twitter, are expanding to attach location data to users' content.

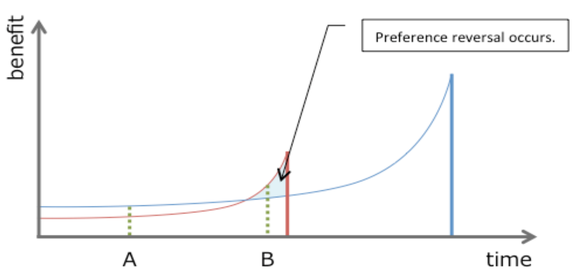


Fig. 1. Preference Reversal Phenomenon

[4] http://foursquare.com/
[5] http://www.google.com/places/

2.2 Filtering Information for Wanderers in Town

In the field of recommender systems, collaborative filtering is one of the popular methods to judge whether information fits the user or not [3][4][5]. The collaborative filtering model is basically based on the assumption that similar users prefer the same information. However, when we consider recommending information to mobile users who are wandering in town, the authors believe that the information must be selected from a set of already filtered candidates in accordance with their situation because the input method and output devices of mobile terminals are highly restricted and also the number of candidates still has to be large even though they are already filtered.

2.3 Phenomena of Human's Preference

In the field of behavioral economics, the phenomena of time preference and temporal discounting are known, which refer to a decrease in the subjective value of a reward as the delay of its receipt increases [6]. People and other animals discount future reward as a function of time. In addition, there is another remarkable phenomenon of preference reversal, which occurs when a subject places a lower selling price on the gamble that he/she chooses than on the other gamble in a pair [7]. For instance, one gamble (the *H* bet) offers a high probability of winning a modest sum of money; the other gamble (the *L* bet) offers a low probability of winning a relatively large amount of money. These bets were also called the *P* bet and the $ bet, respectively, for example,

H bet: 28/36 chance to win $10
L bet: 3/36 chance to win $100

When offered the choice between the two options, most subjects choose the *H* bet over the *L* bet. However, when asked to state their lowest selling price, the majority states a higher price for the *L* bet than for the *H* bet [8]. Therefore, animal and human temporal discounting has been described better as hyperbolic functions than exponential ones in recent psychology.

His notion implies that humans prefer not always rational choices but sometimes irrational and impulsive ones, especially in stressful situations. This paper, therefore, attempts to apply this notion to our model of recommendation for wanderers in town.

3 A Recommendation Model Based on Behavioral Cost

This paper proposes a recommender model, which is based on cost-benefit analysis of the information provided to the user, referring to the model of temporal discounting and preference reversal. Here the authors assume that behavioral cost may be regarded as time in the temporal discounting. Although cost is basically given as distance to reach the location of the information, the authors believe that it depends not only on the geometrical distance but also on his/her cognition. Benefit may be a value that the user can obtain through action influenced by information. The authors would like to emphasize that there can be more preferable information in the middle range than with "low cost–low benefit" or "high cost–high benefit" information if preference reversal phenomena occur. Therefore, a recommender system based on this model can select information, which is located in the surrounding area (not far away) and may be preferred by the user if the system can estimate where the reversal phenomenon occurs.

If a preference reversal phenomenon occurs in the middle range, the system needs to detect its range to determine the preferable information for the user. The authors suppose that the range has to depend on the users' situation, that is, cognitive aspects, such as emotion, objective or destination at that point, accompanying people, physical conditions, such as tiredness, plan of the day, financial status, and so on. To capture such aspects of the users' situation, the system may have to comprehend not only the users' daily activities but also physiological conditions with wearable sensing devices. Of course, it is not easy to perform such daily, lifelong, and extensive logging for humans, even though such technologies are growing rapidly. To begin with, the authors, therefore, try to discover such a range where the reversal phenomenon occurs and to assess the possibility of using such a range for recommender systems.

Fig. 2. pin@clip application. (List mode (left), Map mode (middle), and Information page for Pin (right)).

4 An Experimental System

4.1 pin@clip

The experiments were made on the phenomenon of preference reversals in an experimental social service, called pin@clip, which is an iPhone-based social bookmarking service in Shibuya, Tokyo, Japan that has been operating since December 2009, developed and operated under one of the Japanese governmental projects. Because the pin@clip application[6] for the iPhone is downloadable from Apple's App Store and the service is open and free, the subjects are self-directed.

Using the pin@clip application, users can get "pins", micro blog content of the service. They also can post their own pins, for example, recommendations of their own favorites. When a user posts his/her pin, he or she can choose the location where the pin should be stuck or the store that it should be related to. And also they have to declare their emotion when he/she posts the pin by selecting one of nine emoticons.

[6] pin@clip App. http://itunes.apple.com/jp/app/pin-clip/id338543864?mt=8

Fig. 3. AR mode of pin@clip application

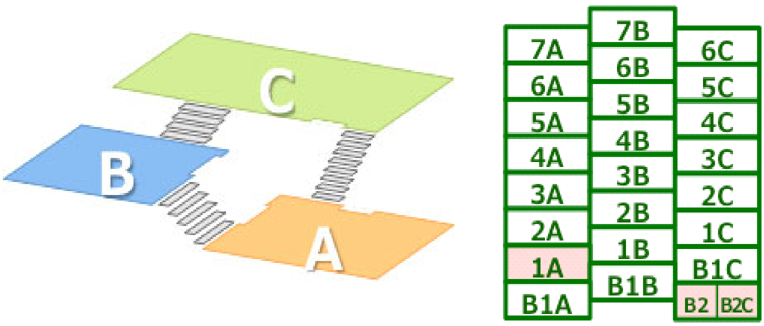


Fig. 4. "Skip floor" architecture of the department store that cooperates in the experiment. Basically each story has three floors. Red shaded floor has entrances.

The pin@clip application provides three display modes: list, map, and augmented reality (AR). The list mode shows pins in a scrollable list. The map mode locates pins in Google Maps. AR mode overlays pins onto the captured image of the iPhone camera in real time.

It also facilitates users to sort pins by five indices in any modes: distance, time line, of similar users, time zones, and recommendation. Time line provides the latest pins in descending order. When the user selects "of similar users", only pins posted by similar users to him/her are selected and sorted by distance. "Time zones" shows pins posted in the same time zone of the current time and sorted by distance. 14 time zones are provided: 12am-6am, 6am-9am, 9am-12pm, 12pm-3pm, 3pm-6pm, 6pm-9pm, and 9pm-12am of weekdays and weekends, respectively. Our recommendation model selects pins for the user when "recommendation" is pushed.

In addition to the basic functions mentioned above, the pin@clip application gives some additional functions for each pin: clip, "give thanks", and "go now". Users can clip their favorite pins for recalling them anytime later. For ratings purposes, they give thanks to preferable pins. If the user decides to go to the location of a pin, he/she sets "go now" to his/her destination.

pin@clip has one special mode for a department store in Shibuya that cooperates in our experiment. When a user enters the store, the pin@clip application shows a dialog box that asks "Do you want to enter the in-store mode?" When he/she enters the in-store mode, the pin@clip application gives not only ordinary pins, such as visitors' pins, but also special pins supplied by store staff. The system could realize his/her location while he/she used the service inside the store, even though GPS signals are unavailable because a significant number of Wi-Fi routers were deployed throughout the building for positioning of the user.

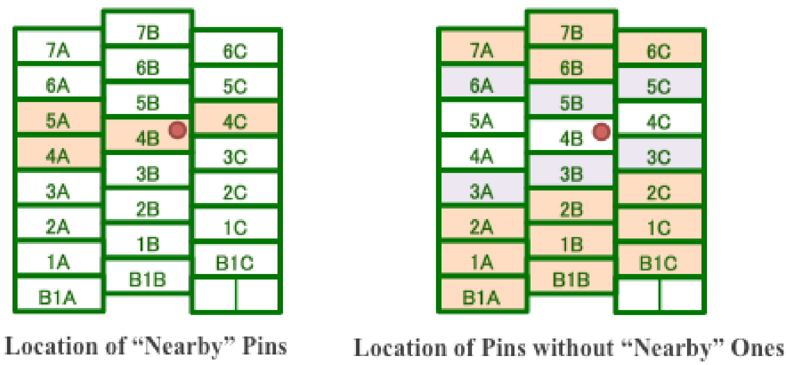


Fig. 5. Locations of Pins to be Provided to the User. (Red dot denotes the location of the user).

4.2 User Data

The pin@clip application collects not only access logs for selected pins, clips, and settings of "go now", but also users' location sensed with GPS and Wi-Fi signals at a specified interval while they use the application. In addition, users are requested to input their gender and age at the first use. Users of the pin@clip service are completely anonymous, and the service doesn't need any personal information, such as name, user account, and email address.

Fig. 6. In-store Mode. (List mode (left) and AR mode (right)).

Similarities and filtering are calculated based on these users' behavior logs. The authors assume that user similarities may depend not only on their demographic attributes and accessed pins that are often used in a lot of existing systems but also on their actual frequency of the accessed area. Because the objective of this paper is not to propose a similarity measure but to apply a behavioral cost model to a recommender system, the authors will explain the similarity measure in detail in another paper.

5 Preliminary Results

5.1 Settings of the Experiment

Place. The experiment was carried out in a department store where users can use the special in-store mode.

Definition of Behavioral Cost. Due to the unique architecture of the building that has seven stories above ground and two basements and each story consists of three "skip" floors (shown in Fig. 4), the authors defined floors between the next upper and lower story as "nearby". Here the authors define the cost to be directly proportional to the number of staircases between two floors.

Group Setting of Users. During the specific period, each user was randomly assigned to one of two groups for comparison when he/she visited a department store for selected products. One was confronted with a list of nearby information; the other was confronted with a list of information without nearby information. 45 subjects were used during the period.

Pins (Provided Content). Prior to the experiment, pins from visitors were prepared, which described a product of the writer's favorite in the floor, and its price was about 1,000 Japanese Yen. The authors assume that benefits of pins can be regarded as equal.

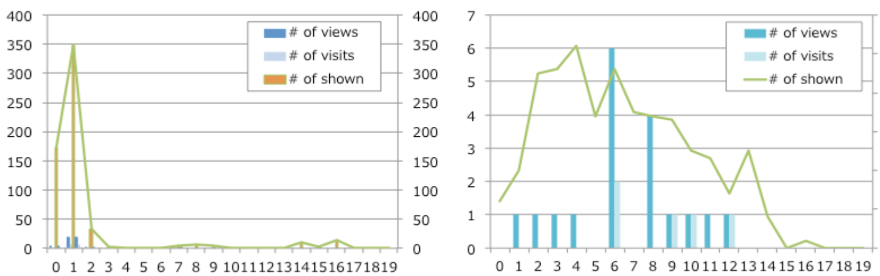
To group A, a list of nearby pins was provided; to group B, a list of pins without nearby ones was provided.

5.2 Results

Fig. 7 shows the experimental results. The graphs include the number of pins shown to the user (right hand y-axis), the number of views by selecting pins in a shown list, and the number of visits to the location of the pin after viewing. The x-axes denote the cost of the shown pin, that is, the number of staircases between the floor where the user is and the floor of the pin. The left graph shows group A results; the right graph shows group B results.

The peak of the number of pins shown to the user is on cost 1 because nearby pins were provided to group A. At cost 14 and 16, provided pins can be observed. The authors consider two possible reasons: one is sensing error and the other is the user accessed the system once and then moved rapidly by using elevators. With group B, it is remarkable that peaks are observed at cost 6 and 8. The authors would like to emphasize that some specific pins that were frequently accessed and preferred were not observed. That suggests that information in this range of cost may be more preferable than nearby and further ones. Although statistically significant data couldn't be obtained in this experiment unfortunately, the observation supports our assumption of the proposed model.

Cost 6 corresponds to 3 stories. Such distance usually brings change of categories of products in the store. Users may be triggered to think of going in another direction when they are confronted with such pins on floors from three or four stories away.



a) Group A

b) Group B

Fig. 7. Experimental Results – Comparison between Groups on Selection of Pins and Actions after Pin Views

6 Conclusion

This paper proposed a new model for recommendation based on behavioral cost of recommendees in town. The model is based on cost-benefit analysis of the information provided to the user, referring to the model of temporal discounting and preference reversal. The experimental results support that a phenomenon of preference reversals might occur, even though the authors could not obtain statistically significant data.

The authors continue to develop and provide the pin@clip service. User logs throughout the area of Shibuya have been taken. Analysis of such logs and the development of methods to capture users' situation including cognitive aspects are future issues.

Acknowledgments. The authors thank Tokyu Corporation, Tokyu Agency Inc., and NEC Corporation for their cooperation with this research. This work is partly supported by the Ministry of Economy, Trade and Industry of Japan.

References

1. Zheng, V.W., Cao, B., Zheng, Y., Xie, X., Yang, Q.: Collaborative Filtering Meets Mobile Recommendation: A User-Centered Approach. In: Proceedings of the 24th AAAI Conference on Artificial Intelligence, pp. 236–241 (2010)
2. Ducheneaut, N., Partridge, K., Huang, Q., Price, B., Roberts, M., Chi, E.H., Bellotti, V., Begole, B.: Collaborative Filtering Is Not Enough? In: Houben, G.-J., McCalla, G., Pianesi, F., Zancanaro, M. (eds.) UMAP 2009. LNCS, vol. 5535, pp. 295–306. Springer, Heidelberg (2009)

3. Su, X., Khoshgoftaar, T.M.: A Survey of Collaborative Filtering Techniques. Advances in Artificial Intelligence, vol. 2009, pp. 1–19 (2009)
4. Schafer, J.B., Frankowski, D., Herlocker, J., Sen, S.: Collaborative Filtering Recommender Systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) Adaptive Web 2007. LNCS, vol. 4321, pp. 291–324. Springer, Heidelberg (2007)
5. Braak, P.t., Abdullah, N., Xu, Y.: Improving the Performance of Collaborative Filtering Recommender Systems through User Profile Clustering. In: Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology, vol. 3, pp. 147–150 (2009)
6. Green, L., Fry, A.F., Myerson, J.: Discounting of Delayed Rewards: A Life-Span Comparison. Psychological Science 5(1), 33–36 (1994)
7. Tversky, A., Kahneman, D.: The Framing of Decisions and the Psychology of Choice. Science 211(4481), 453–458 (1981)
8. Tversky, A., Slovic, P., Kahneman, D.: The Causes of Preference Reversals. American Economic Review 80(1), 204–217 (1990)

A Framework for Agent-Based Simulation in Tourism Planning

Dingding Chao, Kazuo Furuta, and Taro Kanno

Department of Systems Innovation, Graduate School of Engineering,
The University of Tokyo
7-3-1 Hongo, Bunkyo-ku
Tokyo, 113-8656, Japan
{chao,furuta,kanno}@sys.t.u-tokyo.ac.jp

Abstract. Recently, the interest of using agent-based model combined with GIS to perform simulations that seek solutions to problems in the study of tourism services and planning is expanding. However, few scientific studies or systematic methodologies in tourism research have been conducted to support the design and development of such simulations. This research intends to develop a general framework for agent-based simulations in tourism and present its possibility in practical tourism planning process. By developing an agent-based simulation combined with GIS under the protocol, planning supports to tourism bureaus and policy makers to help them assess different tourism policy scenarios and improve tourism services.

Keywords: Tourism services, planning support architecture, agent-based simulation, GIS.

1 Introduction

Recently, the methodology and technique of Agent-Based Simulation (ABS) is applied to increasingly greater number of fields (Drogoul at el., 2005). In tourism, which has become one of the largest industries in many countries (Clancy,1999) and marked by its complex nature, ABS has gained its significance as providing support for the inter-disciplined research in this field (Batty, 2005). As tourism service involves the activities performed by tourists, host communities and the whole industries involved in generating the travel experience (Moutinho, 1987), ABS can be combined with Geographic Information System (GIS) to analyze the spatial-temporal tourism interactions and activities at the macroscopic or microscopic scales.

However, the lack of general principles guiding the design and development of ABS for tourism planning has jeopardized the further integration of the different discipline and approach to cope with the complex nature of tourism services. Agent-based participatory simulation is an approach developed from the integration of Multi Agent Simulation (MAS) and Role-playing Games (RPG) (Guyot and Honiden, 2006) In ecology and resource management research, the joint-use of agent-based model and participatory experiments has been developed through years to deepen modeler's understanding of the interaction between actors involved and to assist decision-making (Bonsquet et al., 1998).

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 280–287, 2011.

Pattern-Oriented Modeling (POM) is an approach proposed by researchers in ecology to describe the real-world complex systems and design agent-based models accordingly (Janssen et al., 2009). With the ultimate goal to overcome the challenges of covering the complexity and eliciting the uncertainty of the system (Grimm et al., 2005), POM approach is used to explore the internal mechanism of the system (Heckbert et al., 2010).We believe that these methodologies can be applied to tourism and provide a scientific strategy for agent-based simulation in tourism planning.

This research intends to present a framework for ABS combined with GIS in the context of tourism service. By integrating POM into the framework, a strategic approach is expected to be proposed to develop ABS for tourism with more reliability. An application example of the framework in the practical developing process of ABS in tourism will be used to illustrate the potential that the framework facilitates researchers and decision makers assessing different tourism policy scenarios.

2 The ABS Framework

In ecology research, Pattern-Oriented Modeling has been used to describe the multilevel human-ecological systems with extracted indicators from observed patterns (Grimm, V., U. Berger, et al., 2006). These indicators are directly related to different spatial and temporal levels to characterize the system inclusively. For tourism, we set up the following steps (Fig.1):

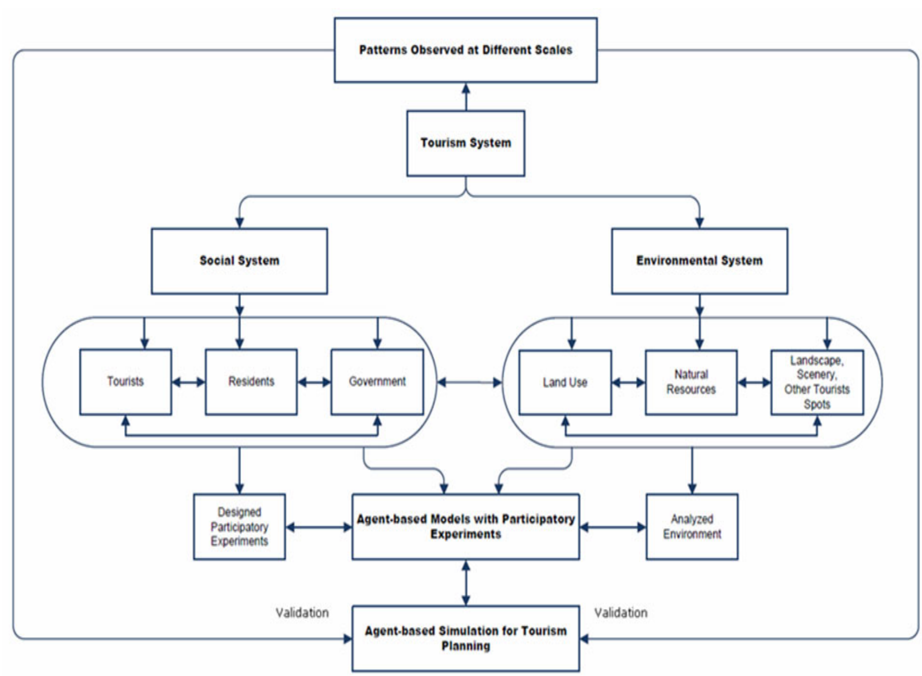


Fig. 1. The Framework

1. Describe the characteristics of tourism systems by observed patterns, set-up sub systems and a spatial-temporal matrix with appropriate spatial and temporal scales for these patterns.
2. Match the indicators of the sub systems which may have great influence on the internal interactions and those emerging patterns at macroscopic or microscopic level onto the spatial-temporal matrix accordingly.
3. Establish hypothesis, models, simulation environments based on the indicators.
4. Develop Agent-based Simulation based on the parameter and proposed internal dynamics perform participatory experiments to replace some of the artificial agents. If ABS is combined with GIS, parameters can be linked to the specific layer containing corresponding data processed in GIS
5. Perform simulation and compare simulated results to the observed pattern to modify the present model or choose an alternative model.

3 Application Sample : An Agent-Based Simulation for Tourisn Planning

3.1 Overview

Our SimHakone model (Chao, D., Furuta, K., and Kannon, T., 2010) for tourism planning in Hakone area is based on some works using agent-based models to study land use-land use change (Parker, et al., 2002; Batty, 2005; Li and Liu, 2008) with modifications regarding the differences between the characteristics of urban development system and tourism system.

The procure of developing the system follows the proposed frame work will be described in the following chapters.

3.2 Spacial and Temporal Scales Setup

Tourism plannning invovles planning activities at various hirechical levels (The World Commission on Environment and Development Publication, 1994) so that apporpriate spacial and temporal scales need to be defined. The following table includes the possible spatial and temporal scales of the observed patterns.

Table 1

<table>
<tr><th>Spatial Scales</th><th>Temporal Scales</th></tr>
<tr><td>Attractions</td><td>Hours</td></tr>
<tr><td rowspan="2">Tourism Services Districts</td><td>Days</td></tr>
<tr><td>Weeks</td></tr>
<tr><td>Regions</td><td>Months</td></tr>
</table>

Table 1. (*Continued*)

Countries	Seasons
Global	Years
	Decades
	Centuries

3.3 Definition of Patterns and Variables

In order to describe the behavior of a tourism system and to develop agent-based models, appropriate indicators must be extracted from the emerging patterns, which can generally be divided into four groups: economic, environmental, social, and multi-dimensional. The indictors were re-projected also according to their temporal and spatial scales into the matrix as shown in Figure 2 (Adapted from Becker, 1999).

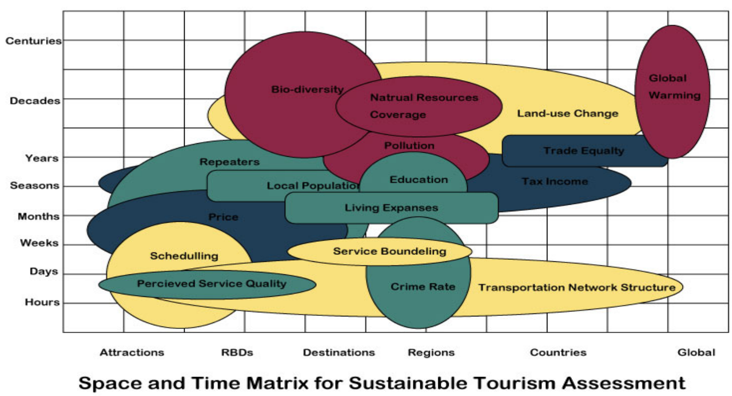


Fig. 2. Sustainable Tourism Indicators (Adapted from Becker, 1999)

3.4 GIS Layers

Spatial and statistical data can be processed using GIS and distributed into several layers of the model to represent the charateristics of the complex system that the agents share and interact with. Essentially, there are the following layers: Land Type Layer, Transportation Layer, Tourist Spot and Public Facility Layer, and Tourists and Residents Layer. Reprojection on to a ABS platform is ususally needed for future simulation. Figure.3 illustrate an example of using ArcGIS (ESRI) as the GIS to process data and Netlogo (Wilensky, U. 1999) as the simulation platform.

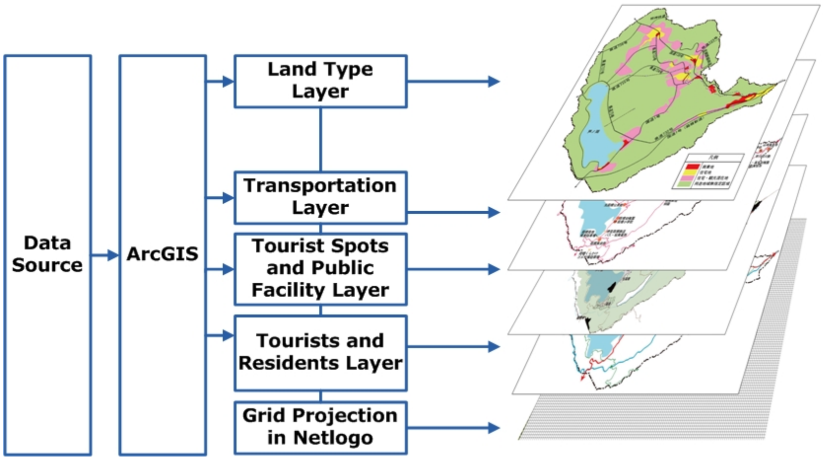


Fig. 3. Simulation Structure

3.5 SimHakone Interface

The interface of SimHakone consists of the following parts (Figure):
Main Control: Setup the original map and On/Off control
Scenario Picker: Choose one of the development scenarios for simulation
Variables Set-up: Can adjust different variables for simulation
Output Control: Report the results data to a .csv file
Viewer: View the process of the simulation
Plot: Plot the result data

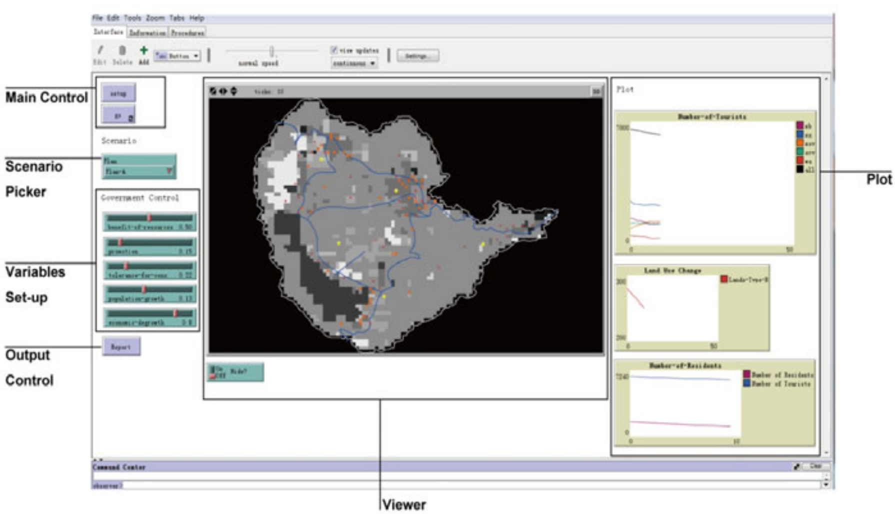


Fig. 4. Interface

3.6 Simulation and Results

A case study of test simulation was performed to demonstrate the usefulness of the proposed ABS for sustainable tourism development planning. Simulation scenarios Both the macroscopic and the microscopic level model are include in the simulation

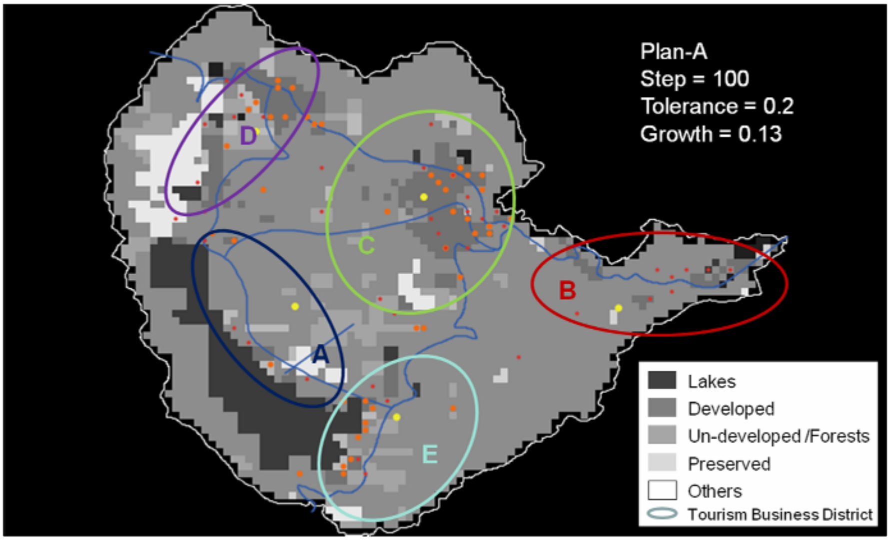


Fig. 5. Simulation Case Study

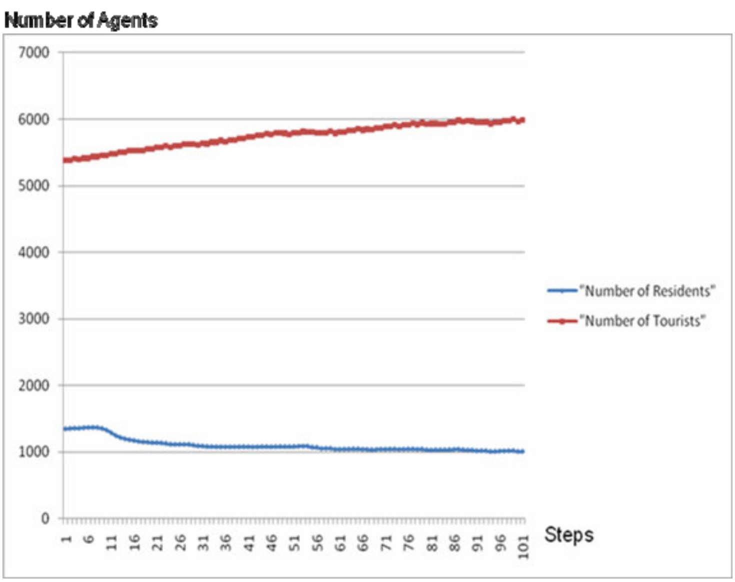


Fig. 6. Number of agents resulted from Plan A

for examining the planning scenarios. The indicators discussed in the previous chapter are included for assessing the consequences of the planning.

We divided the study area into 5 major RDBs (marked with A, B, C, D, E in Fig.5) to observe the change in land use and number of tourists in each district. It is revealed in the Simulation that the changes in government promotion (increase of the ratio of tourists to residents) which results in changes in tourists flow among different RBDs will largely influence their development pattern. FIGURE 6 shows the simulated result of land use change after taking one of the plans. The changes in land use and numbers of tourists can be reflected by the simulation and provide decision supports to the government.

4 Conclusion

Systematic methodology is needed for rational planning and sustainable development of tourism services, but the human-environmental system is complex and requires combination of studies in various disciplines. Agent-based modeling offers an approach to integrate interdisciplinary studies and simulate the complex processes for facilitating tourism services planning. This article presented a framework for designing and developing ABS for tourism planning. Variables are coupled with specific special-temporal scales to indentify patterns at both macroscopic and microscopic level. We confronted the challenge of the complex nature of tourism services by designing the simulation model under the framework. The sample ABS simulation developed under the framework demonstrated the potential that ABS can be applied into the practical planning and policy making process for Tourism Bureaus.

References

Batty, M.: Cities and Complexity: Understanding Cities with Cellular Automata, Agent-Based Models, and Fractals. MIT Press, Cambridge (2005)

Becker, B.: Sustainability assessment: a review of values, concepts, and methodological approaches. Consultative Group on International Agricultural Research, The World Bank, Washington, D.C., USA (1997)

Bonsquet, F.: CORMAS: Common pool resources and multi-agent system. Lecture Notes in Computer System 14(16), 826–838 (1998)

Butler, R.: The concept of a tourism area cycle of evolution: Implications for resources. Canadian Geographer 24(1), 5–12 (1980)

Carter, R.W., Baxter, G.S., Hockings, M.: Resource management in tourism research: A new direction? Journal of Sustainable Tourism 9(4), 265–280 (2001)

Chao, D., Furuta, K., Kannon, T.: Agent-Based Simulation System for Supporting Sustainable Tourism Planning. Submission to Post Proceedings of JSAI International Workshop (2010)

Clancy, M.: Tourism and development: Evidence from Mexico. Annals of Tourism Research 26, 1–20 (1999)

Tisdell, C.A.: Tourism Economics, the Environment and Development: Analysis and Policy. Edward Elgar Publishing (2001)

Drogoul, A., Vanbergue, D., Meurisse, T.: Multi-agent Based Simulation: Where Are the Agents? In: Sichman, J.S., Bousquet, F., Davidsson, P. (eds.) MABS 2002. LNCS (LNAI), vol. 2581, pp. 1–15. Springer, Heidelberg (2003)

ESRI: ArcGIS Desktop 9.3 (2008), http://www.esri.com/
Graham, S., Steiner, J.: TravellerSim: Settlements, Territories and Social Networks (2006), http://home.cc.umanitoba.ca/~grahams/Travellersim.html
Grimm, V.: Pattern-Oriented Modeling of Agent-Based Complex Systems: Lessons from Ecology. Science 310(5750), 987–991 (2005)
Grimm, V., Berger, U., et al.: A standard protocol for describing individual-based and agent-based models. Ecological Modelling 198(1-2), 115–126 (2006)
Guyot, P., Honiden, S.: Agent-Based Participatory Simulations: Merging Multi-Agent Systems and Role-Playing Games (2006)
Heckbert, S., Baynes, T., Reeson, A.: Agent-based modeling in ecological economics. Annals of the New York Academy of Sciences 1185, 39–53 (2010)
Janssen, M.A., Radtke, N.P., Lee, A.:: Pattern-Oriented Modeling of Commons Dilemma Experiments. Adaptive Behavior 17(6) (2009)
Li, X., Liu, X.: Embedding sustainable development strategies in agent-based models for use as a planning tool. International Journal of Geographical Information Science 22(1), 21–45 (2008)
Moutinho, L.: Consumer Behaviour in Tourism. European Journal of Marketing 21(10), 5–44 (1987)
Malaka, R., Zipf, A.: A, DEEP MAP -Challenging IT research in the framework of a tourist information system. In: Fesenmaier, D., Klein, S., Buhalis, D. (eds.) Information and Communication Technologies in Tourism 2000. Proceedings of ENTER 2000, pp. 15–27. Springer Computer Science, Wien (2000)
Parker, D.C., et al.: Multi-Agent Systems for the Simulation of Land use and Land-Cover Change: A Review. In: Special Workshop on Agent-Based Models of Land use, Irvine, California, USA, October 4-7 (2002)
The World Commission on Environment and Development Publication, National and Regional Tourism Planning: Methodologies and Case Studies (1994), http://pub.unwto.org/WebRoot/Store/Shops/Infoshop/Products/1015/1015-1.pdf
Wilensky, U.: NetLogo, Center for Connected Learning and Computer-Based Modeling. Northwestern University, Evanston, IL (1999), http://ccl.northwestern.edu/netlogo
World Tourism Organization: Tourism Vision (2020), http://unwto.org/facts/eng/vision.htm
Wu, F.: Calibration of stochastic cellular automata: the application to rural–urban land conversions. International Journal of Geographical Information Science 16, 795–818 (2002)

Safe-in-Place Awareness GPS System with Distance-Based and Duration-Based Notification Control

Chi Nung Chu and Gene Chu

China University of Technology, Department of Management of Information System, No. 56, Sec. 3, Shinglung Rd., Wenshan Chiu, Taipei, Taiwan 116, R.O.C.
nung@cute.edu.tw

Abstract. This paper describes the design of SIPGPS (Safe-in-Place GPS System) which is a GPS assistance environment helping track the elderly people outdoors. This environment is intended to facilitate the use of the GPS cell phones in assisting elderly people with walking safely in daily life. SIPGPS facilitates safe-in-place by assisting elderly people with emergency assistance via distance-based and duration-based notification control. The SIPGPS which provides the dedicated family easier way to locate their elderly family may expand the role of care in remote location.

Keywords: Safe-in-Place GPS System**,** Participatory Design.

1 Introduction

GPS cell phones are emerging as a new opportunity for the elderly people walking safe in place. The development of cell phones with embedded GPS engines has been a reality in many platforms of smart phones, such as iPhone, Google Android phone, and Windows phone. These phones, in part, are mandated requiring the position of a cell phone to be available to emergency call dispatchers, but the lack of self-conscious awareness for the elderly people means that many are missing out or positioning a critical condition with or without such critical service.

There are three basic types of technologies GPS, GSM and GIS incorporated with cars to provide tracking information. Thus the assistance server in security center is able to compute position solutions, leaving the GPS receiver with the sole job of collecting range measurements. As smart home technology emerges, it could be augment its usage to facilitate aging-in-place by assisting patients with emergency assistance, fall prevention/detection, reminder systems, medication administration and assistance for those with hearing, visual or cognitive impairments.

This paper intends to make aging-in-place a reality, continuous monitoring, and improved psychosocial effects. The concept of safe-in-place awareness expands the role of the caring in the future. It is important for all family to understand how their concerns will be transformed as remote-caring become a reality for the aging population.

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 288–292, 2011.

2 Safe-in-Place GPS System

Safe-in-Place GPS System (SIPGPS) describes a system where outside sources, such as a monitoring server (Fig. 1) and instant message network, and an inside GPS receiver agent (Fig. 2), help a GPS cell phone perform the tasks required to make range measurement and duration of stay at a place measurement for alert notification. The GPS receiver agent installed in GPS cell phone could connect to monitoring server every 60 seconds with collected cell phone numbers from the SIM chip and longitude and latitude from the GPS receiver of cell phone. The monitoring server communicates with GPS cell phone via a wireless link and measures those information to trace

Fig. 1. Homepage of Monitor Server

Fig. 2. GPS receiver agent installed in GPS cell phone

out the distance moved and time stayed in a Google map dynamically (Fig. 3). According to the dedicated family managed safe range and duration of activity in the monitoring server, they could log on to identify where the elderly people are in the safe range of area and if they stayed still at one position for some potential health reasons. As the alerts of violated distance or duration triggered, the dedicated family could receive email, instant message and text message through phone and computer. Therefore the resulting SIPGPS system could boost performance beyond that of the GPS cell phone in a stand-alone mode.

Fig. 3. Tracking measurements of safe range and duration of stay

3 Evaluation of Safe-in-Place GPS System

The focus of SIPGPS has been to design an online interaction and community that would increase family's motivation, commitment and satisfaction with the online care system. The Participatory Design methodology originating from Scandinavian software development traditions [3] encouraged the participation of effected workers in technology development processes [6]. Participatory Design advocates the user involvement is seen as critical both because users are the experts in the work practices supported by the technologies and because users ultimately will be the ones creating new practices in response to new technologies [1, 2, 4, 5]. Engaging participation of potential end-users and stakeholders in open but monitored environments allows for exploration of design responses to situations generated by daily lives. In Participatory Design, the prototype of SIPGPS is central to the notion of design-by-doing that is a necessary component for simulating situated environments as natural experiments [7]. In the case of design research, combined uses of prototype exist that are themselves outcomes and conditions for experimental research.

This experiment used the Participatory Design to evaluate the efficiency and effectiveness of supports by SIPGPS. SIPGPS consists of two main parts, monitoring server and instant message network used by dedicated family, and a GPS receiver agent used by subjects' cell phones. 6 subjects of elderly people with dedicated family were conducted in this study. Each subject's dedicated family members could use the SIPGPS. Digital video cameras held by researchers were used to capture each subject's movements. Subjects were all told their daily outside activities would be recorded without their notice to avoid interfering with their natural behaviors. During the experiment, researchers may engage in an active discussion of what the participants are doing, or have them think aloud along the way. Researchers may also consider waiting until participants are finished, and ask questions at that point about the functions they have chosen. This all depends on researchers' goals for the study.

The target event is comprised of monitored subjects who walk out of family-defined distance in potential dangerous area around subject's residence (such as busy transportation zone), monitored subjects who stay still somewhere over some setup time (such as heart attack), or monitored subjects who are more commonly in need of assistance (such as stray or health problem). These monitored subjects benefit from being able to be located somewhere immediately from a remote location by the dedicated family and requesting assistance by pressing a single button. These actions are made possible by exact location obtained by using the GPS receiver agent with SIPGPS.

3.1 Results and Discussion

Participatory Design in the study elicited ideas that the subjects needed awareness too while some event occurred (e.g. cell phone vibration), multi-response to dedicated family (e.g. text message and email), and more interaction interface other than visual clues (e.g. voice).

The evaluation results of SIPGPS are shown in Table 1. Dedicated family members received notification significantly fast enough in 2 minutes when the safe distance or period of staying still was measured. After issuing any emergency call, the subject could be calmed down in 7 minutes by getting a solution message from the family and waited for the actual rescue. These data meant that safe awareness has been adapted to family concerns in that elderly people could be taken care of in time with SIPGPS.

Table 1. Efficiency Evaluation of Safe-in-Place GPS System (minutes)

Event	Response Time*	Waiting Time**
Out of family-defined distance	1.6	2.2
Staying still	2.1	10.8
Requesting assistance	3.5	5.8

* Response time – amount of time it takes from when a event was submitted until the first response is produced

** Waiting time – amount of time a request has been waiting in the assurance of assistance solution, not assistance acquisition.

SIPGPS is most useful at a situation as the dedicated family is concerned about the subject's certain conditions if he/she walks alone in a safe state. The more concerns bring the more watching on SIPGPS. The event response time in the monitoring context reveals the efficiency of SIPGPS. While event "staying still" occurs, waiting time shows the complicated caring process in finding out what is going on for the elderly people with no movement, such as falling down, heart attack or chatting with some. The SIPGPS tends to elicit effective remote caring from the Internet.

4 Conclusions

Mobile services play an important role in assisting elderly people with lack of self-conscious awareness in daily life. SIPGPS facilitates safe-in-place by assisting elderly people with emergency assistance via distance-based and duration-based notification control. The SIPGPS which provides the dedicated family easier way to locate their elderly family may expand the role of care in the future. Benefits include making safe-in-place a reality, continuous monitoring, and improved the well being of aging society.

References

1. Baroudi, O., Ives: An Empirical Study of the Impact of User Involvement on System Usage and Information Satisfaction. CACM 29(3), 232–238 (1986)
2. Belotti, V., Bly, S.: Walking away from Desktop Computer: Distributed Collaboration and Mobility in a Product Design Team. In: Proceedings of CSCW 1996, pp. 209–218. ACM Press, Cambridge (November 1996)
3. Bjerknes, G., Ehn, P., Kyng, M. (eds.): Computers and Democracy - A Scandinavian Challenge. Aldershot, England (1987)
4. Blomberg, J.L., Henderson, A.: Reflections on Participatory Design: Lessons from the Trillium Experience. In: Proceedings of CHI 1990, pp. 35–359. ACM Press, Seattle (1990)
5. Button, G., Sharrock, W.: Project Work: the Organisation of Collaborative Design and Development in Software Engineering. CSCW Journal 5(4), 369–386 (1996)
6. Floyd, C., Mehl, W.-M., Resin, F.-M., Schmidt, G., Wolf, G.: Out of Scandinavia: Alternative Approaches to Software Design and System Development. Human-Computer Interaction 4(4), 253–350 (1989)
7. Hollan, J., Hutchins, E., Kirsh, D.: Distributed Cognition: Toward a New Foundation for Human-Computer Interaction Research. ACM Transactions on Computer–Human Interaction 7(2), 174–196 (2000)

Landmarks Detection to Assist the Navigation of Visually Impaired People

Paulo Costa[1], Hugo Fernandes[2], Verónica Vasconcelos[3], Paulo Coelho[1], João Barroso[2,4], and Leontios Hadjileontiadis[5]

[1] Polytechnic Institute of Leiria, Portugal
paulo.costa@ipleiria.pt, paulo.coelho@ipleiria.pt
[2] University of Trás-os-Montes e Alto Douro, Portugal
[3] Polytechnic Institute of Coimbra, Portugal
[4] GECAD – Knowledge Engineering and Decision Support Research Center, Porto, Portugal
[5] Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Greece
{Hugof,jbarroso}@utad.pt, veronica@isec.pt, leontios@auth.gr

Abstract. Assistive technology enables people to achieve independence in the accomplishment of their daily tasks and enhance their quality of life. Visual information is the basis for most navigational tasks, so visually impaired individuals are at disadvantage due to the lack of information or given insufficient information about their surrounding environment. With the recent advances in inclusive technology it is possible to extend the support given to people with visual disabilities during their mobility. In this context we propose a new algorithm to recognize landmarks suitably placed on sidewalks. The proposed algorithm uses a combination of Peano-Hilbert Space Filling Curves for dimension reduction of image data and Ensemble Empirical Mode Decomposition (EEMD) to pre-process the image, resulting on a fast and efficient recognition method and revealing a promising solution.

1 Introduction

Assistive technology enables people with disabilities to accomplish daily tasks and assists them in communication, education, work and recreational activities. Principally though, it can help them to achieve greater independence and enhance their quality of life. From the various assistive technologies available nowadays, a special focus was put on those that help blind or visually impaired people with their mobility. The World Health Organization estimates that there are 37 million blind people worldwide [1]. Blind or visually impaired people have a considerable disadvantage, as they need information for bypassing obstacles and have relatively little information about landmarks, heading and self-velocity. The main issue on using assistive technologies is to provide additional information, useful to blind people during their mobility process, i.e., walking.

Human mobility can be distinguished between Orientation and Navigation. Orientation can be thought of as knowledge of the basic spatial relationships between

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 293–300, 2011.

objects within the environment. Information about position, direction, desired location, route, route planning, etc., are all bound up with the concept of orientation. Navigation, in contrast, suggests an ability to move within the local environment. This navigation implies the knowledge of immediate objects and obstacles, ground morphology (holes, stairs, flooring etc.), and moving or stationary dangers.

The aim of the present work is to propose a new algorithm for the computer vision (CV) module that will be integrated in a prototype currently being developed by the SmartVision project team. The new image processing algorithm is intended to extract useful information from outdoor scenes in the University of Trás-os-Montes and Alto Douro (UTAD) campus and put the blind user correctly positioned on the sidewalk along a predefined route. In order to reduce the image complexity features for extraction/detection, some landmarks were placed along the sidewalks. The proposed algorithm uses a combination of Peano-Hilbert space filling curves [2] and Ensemble Empirical mode decomposition (EEMD) [3] for image processing and a basic correlation algorithm for template matching.

The paper is organized as follows. Section 2 presents a classification of navigation systems and related work; some projects that represent the state of the art are presented. Section 3 presents the proposed algorithm and the related techniques used. Section 4 presents and discusses the results. Finally, Section 5 concludes the paper.

2 Background and Related Work

An Electronic Travel Assistant (ETA) has to supply the visually impaired with the necessary routing information to overcome obstacles in the near environment with minimum error. Navigation systems to assist visually impaired people can be classified in three groups, based on their usage. The indoor systems are to be used in structured environments with less complex scenes, typically inside buildings or in isolated controlled campuses. The outdoor navigation systems are intended to be used in exterior open space, typically on the street. The indoor/outdoor systems can be used in both indoor and outdoor spaces, switching functionalities based on environment operation.

Some commercial research and development (R&D) projects that currently describe the state of the art in outdoor navigation systems for assisting visually impaired people are presented as follows:

1. Navigation systems without local obstacle information: the systems BrailleNote GPS [4], StreetTalk [5], Trekker [6], NOPPA [7], Navigator [8] and Drishti [9] are GPS based systems to assist the navigation of visually impaired people. Their primary components are a PDA or Laptop especially designed/adapted for people with visual disabilities, a Bluetooth GPS receiver and specially developed software for configuration, orientation and route mapping. The output give for user interaction can be a Braille display or a speech synthesizer.
2. Navigation systems with local obstacle information provide better knowledge of the local scenario, increasing the information quality provided to the blind user to overcome local obstacles.

Several techniques are used to detect and measure object distances, like multiple ultrasonic sensors (sonar) [10] and Laser Range Scanner (LRS) [11], and computer vision (CV) techniques like Principal Component Analysis (PCA) used in ASMONC [10], the Fuzzy Like Reasoning segmentation technique used in the Tyflos system [11], the Expectation-Maximization (EM) algorithm used by Zelek [12], the stereo images for measuring distance to objects, used by Meers [13] and Hadjileontiadis [14], the Neural Network technique used in NAVI [15] and, later on in the same project, the authors also tested Fuzzy Learning Vector Quantification (FLVQ) to classify objects in the scene.

3 Peano-Hilbert and Ensemble Empirical Mode Decomposition

Empirical Mode Decomposition (EMD) [16] is a method for breaking down the signal without leaving the time domain; it filters out functions which form a complete and nearly orthogonal basis for the signal being analyzed. The adoption of adaptive basis functions introduced by Huang et al. [16] provided the means for creating intrinsic *a posteriori* base functions with meaningful instantaneous frequency in the form of Hilbert spectrum expansion [16]. These functions, known as Intrinsic Mode Functions (IMFs), are sufficient to describe the signal, even though they are not necessarily orthogonal [16]. IMFs, computed via an iterative 'sifting process' (SP), are functions with zero local mean [16], having symmetric upper and lower envelopes. The SP depends both on an interpolation method and on a stopping criterion that ends the procedure. Some updates of the 1D-EMD have been proposed which address the mode mixing effect that sometimes occurs in the EMD domain. In this vein, 1D-Ensemble EMD (1D-EEMD) has been proposed [3], where the objective is to obtain a mean ensemble of IMFs with mixed mode cancelation due to input signal noise addition.

3.1 1D-Empirical Mode Decomposition (1D-EMD)

1D-EMD considers a signal $x(t)$ at the scale of its local oscillations [16]. Locally, under the EMD concept, the signal $x(t)$ is assumed as the sum of fast oscillations superimposed to slow oscillations. On each decomposition step of the EMD, the upper and lower envelopes are initially unknown; thus, an interactive sifting process is applied for their approximation to obtain the IMFs and the residue. The 1D-EMD scheme is realized according to the following steps [16]:

1. Identify the successive extrema of $x(t)$ based on the sign alterations across the derivative of $x(t)$;
2. Extract the upper and lower envelopes by interpolation; that is, the local maxima (minima) are connected by a cubic spline interpolation to produce the upper (lower) envelope. These envelopes should cover all the data between them;
3. Compute the average of upper and lower envelopes, $m_1(t)$;
4. Calculate the first component $h_1(t) = x(t) - m_1(t)$;
5. Ideally, $h_1(t)$ should be an IMF. In reality, however, overshoots and undershoots are common, which also generate new extrema or exaggerate the existing ones

[16]. To correct this, the sifting process has to be repeated as many times as is required to reduce the extracted signal as an IMF. To this end, $h_1(t)$ is treated as a new set of data, and steps 1-4 are repeated up to k times (e.g., $k = 7$) until $h_{1k}(t)$ becomes a true IMF. Then set $c_1(t) = h_{1k}(t)$. Overall, $c_1(t)$ should contain the finest scale or the shortest period component of the signal;

6. Obtain the residue $r_1(t) = x(t) - c_1(t)$;
7. Treat $r_1(t)$ as a new set of data and repeat steps 1-6 up to N times until the residue $r_N(t)$ becomes a constant, a monotonic function, or a function with only one cycle from which no more IMFs can be extracted. Note that even for data with zero mean, $r_N(t)$ still can differ from zero;
8. Finally,

$$x(t) = \sum_{i=1}^{N} c_i(t) + r_N(t), \tag{1}$$

where $c_i(t)$ is the -th IMF and $r_N(t)$ the final residue.

3.2 1D-Ensemble Empirical Mode Decomposition (1D-EEMD)

One of the major drawbacks of the original 1D-EMD is the appearance of mode mixing, which is defined as a single IMF consisting of signals of widely disparate scales, or a signal of similar scale residing in different IMF components. The effect of adding white noise scales uniformly through the whole time-scale or time-frequency space, will provide a reference distribution to facilitate the decomposition method. The added white noise may also help to extract the true signals in the data, a truly Noise-Assisted Data Analysis [3]. The 1D-EEMD is implemented as follows:

1. Add white noise series w(t) to the data x(t), X(t) = x(t) + w(t);
2. Decompose the X(t) data with white noise into IMFs, $X(t) = \sum_{j=1}^{N} c_j(t) + r_N(t)$;
3. Repeat step 1 and step 2 several times with different noise series $w_i(t)$, $X_i(t) = x(t) + w_i(t)$, and obtain corresponding IMFs, $X_i(t) = \sum_{j=1}^{N} c_{ij}(t) + r_{iN}(t)$;
4. Finally, the ensemble means of corresponding IMFs of the decomposition are

$$c_j(t) = \frac{1}{N} \sum_{i=1}^{N} c_{ij}(t), \tag{2}$$

where N is the ensemble members.

3.3 Peano Hilbert-Ensemble Empirical Mode Decomposition (PH-EEMD)

A space-filing curve (SFC) is a continuous scan that passes through every pixel of the image only once. In order to transform an image (2D data) on a signal (1D), the space filing curve must preserve the neighborhood properties of the pixel [17]. These curves were first studied by Peano [18] and later by Hilbert [19] and corresponding algorithms are described in [2]. The Peano-Hilbert curve has three main interesting properties: (i) the curve is continuous; (ii) a scanning curve is continuous almost everywhere; and (iii) some parts of the curve are similar with whole curve suggesting a fractal structure. The Peano-Hilbert curve is the most popular recursive SFC and is used in many applications. Fig.1 represents three Peano-Hilbert space filing square curves of area one.

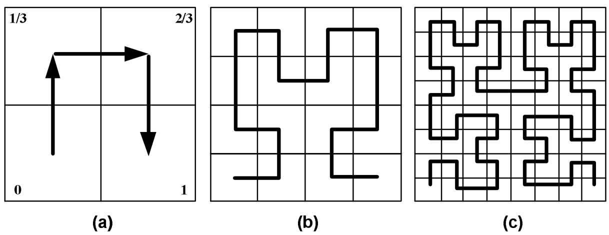


Fig. 1. Peano-Hilbert Curve: a) basic curve; b) 2 interactions; c) 3 interactions

The advantages of space filing curves were combined with the 1D-EEMD algorithm proposed by Huang for image processing. Some extensions to 2D-EMD were proposed to deal with the 2D nature of the data, but these fully 2D-EMD approaches are very time consuming processes. An algorithm for 2D application was proposed with relatively low trend in processing time. This algorithm is based on three phases:

1. Decompose the image using the Peano-Hilbert curve and get the equivalent 1D signal. For the Peano-Hilbert algorithm, a recursive function is used to get the -th order area one curve.

This procedure converts 2D data into 1D signal maintaining the local pixel spatial relations between neighbors.

2. Apply the 1D Ensemble Empirical Mode Decomposition (EEMD) to the linear signal in order to compute the 1D Intrinsic Mode Functions that carry multi-scale space-frequency information. For the EEMD white noise with amplitude of 0.1, standard deviation of the original data is added and the process is repeated 8 times.

Boundaries problems are associated to most of the data processing algorithms due to finite data samples. In the EMD algorithm this is particularly true in the interpolation procedure on the sifting process; to solve this, an even data extension method is used [16].

3. To get the 2D data decomposition the inverse procedure must be taken to reconstruct the image from the data, using the Peano-Hilbert pixel spatial relations to process the 1D IMFs back to 2D IMFs, according Fig.1.

3.4 The Proposed Algorithm for Landmarks Detection

In order to provide useful information to blind people the vision system must be able to detect relevant features in the scene and help the blind user to stay in safe paths. The first approach was intended to reduce the image complexity for the processing algorithms and to enhance the detection. Landmarks were made on sidewalks representing safe paths along the route. From several geometric marks, circles were adopted because in this application they are scale and rotational invariant. Several captured images with different circle radius were tested and to minimize the size and maximize de detection rate a 15 cm circle radius was chosen.

The proposed CV phases are described below:

1. Decompose the captured image with Peano-Hilbert Ensemble Empirical Mode Decomposition
2. Image filtering to eliminate higher frequencies containing noise and fine details. This process is achieved in the EEMD reconstruction phase (1) by eliminating the first two IMFs according to a criterion of root mean square error (RMSE) minimization. Fig.2. represents the *RMSE* during the reconstruction phase of the image which was corrupted by Gaussian white noise ($\sigma = 0.1$). The minimum error occurs when removing the first two IMFs, where the reconstruction is $x(t) = \sum_{i=3}^{N} c_i(t) + r_N(t),$ were i starts at 3 IMF.

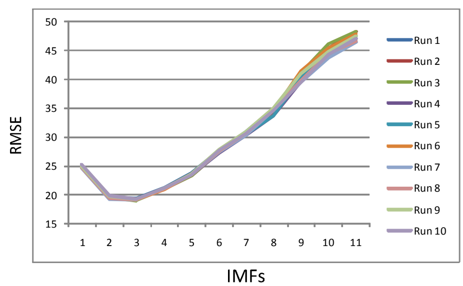


Fig. 2. Reconstruction error of a image corrupted with white noise

3. Define a region of interest (ROI) near the blind user; in our case we chose to analyze the first half of the image and a quarter image size for each side of the blind user. Data outside the ROI area is set to zero.
4. Perform data binarization of the ROI image with a global threshold using Otsu's method [20], followed by Canny edge detection.
5. Finally the ROI image is passed onto a circle detection procedure using simple correlation template matching.

4 Experimental Results and Discussion

In order to test the proposed algorithm for the detection of landmarks within the sidewalk, a set of different images were captured on UTAD campus (outdoor

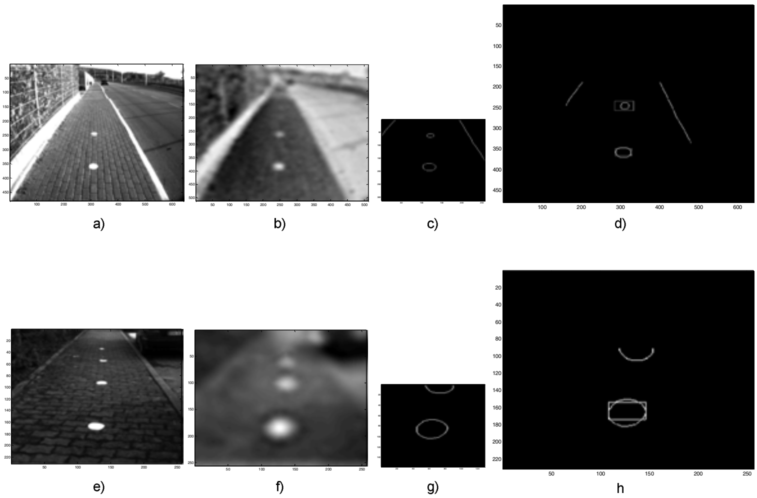


Fig. 3. Results of the proposed algorithm for landmarks detection

scenario). Two images that represent different sections of the campus are depicted in Figs. 3a) and 3e), with the latter representing a more difficult task for any image processing algorithm. Fig. 3b) and Fig. 3f) are the corresponding EEMD filtered images. As it can be seen from these figures the higher frequencies were removed and, with this procedure of prior image binarization, the appearance of small artifacts was minimized. Fig. 3c) and Fig. 3g) show the ROI near the user that are processed in circle detection. We consider that the user is centered at the bottom of image. Finally, Fig. 3d) and Fig. 3h) represent the circle detection results of the respective images; all detected circles are marked with a rectangle. In order to improve the visualization of the results, these two images are presented at a bigger scale.

From the outputs of circle detection in the image, the blind user must head in the direction of the nearest circle. This ensures that s/he will not get out of safe path. Based on the relative position of the nearest circle to the blind user it is possible to compute the trajectory correction and output it to the blind user. The interface to the user is made through five microvibrators, corresponding to five directions, i.e., left, left-diagonally, straight, right-diagonally and right.

5 Conclusion and Future Work

In the presented work a Computer Vision module for the SmartVision project was proposed. For an efficient assistance to a blind user's navigation the CV module must detect accurately specific features in the environment. Due to very different scenarios that can be found in outdoor navigation the sidewalks were marked with landmarks to improve the CV feature detection efficiency. For landmark detection the Peano-Hilbert Ensemble Empirical Mode Decomposition Template Matching method was implemented and the system has proved to be able to detect the defined landmarks and provide valid and simple instructions to the blind user.

Further work is needed to enhance the method accuracy and future improvements will still continue to use PH-EEMD image analysis. Range image information (disparity map) will be integrated into the CV module to provide obstacle detection features.

References

1. WHO: Magnitude and causes of Visual Impairment. Fact sheet no. 282 (November 2004)
2. Allaart, P.C., Kawamura, K.: Dimensions of the coordinate functions of space-filling curves. Journal of Mathematical Analysis and Applications 335, 1161–1176 (2007)
3. Wu, Z., Huang, N.E.: Ensemble empirical mode decomposition: a noise-assisted data analysis method. Centre for Ocean-Land-Atmosphere Studies, Calverton, Md, USA (2005)
4. GPS, `http://www.senderogroup.com` (last viewed on Janury 2011)
5. StreetTalk, `http://www.freedomscientific.com` (last viewed on January 2011)
6. Trekker, `http://www.humanware.ca` (last viewed on Janury, (2011)
7. Ari, V., Sami, K.: NOPPA - Navigation and Guidance System for the Blind. VTT Industrial Systems, ITS World 2004 (2004)
8. Kowalik, R., Kwasniewski, S.: Navigator – A Talking GPS Receiver for the Blind. In: Miesenberger, K., Klaus, J., Zagler, W.L., Burger, D. (eds.) ICCHP 2004. LNCS, vol. 3118, pp. 446–449. Springer, Heidelberg (2004)
9. Ran, L., Helal, S., Moore, S.: Drishti: An Integrated Indoor/Outdoor Blind Navigation System and Service. In: Proceedings of IEEE International Conference on Pervasive Computing and Communication 2004, Orlando, pp. 23–31 (2004)
10. Molton, N., Se, S., Brady, J.M., Lee, D., Probert, P.: A stereo vision-based aid for the visually impaired. Image and Vision Computing 16, 251–263 (1998), 0262-8856/98/$19.00 q 1998 Elsevier Science B.V. All rights reserved
11. Bourbakis, N.G., Kavraki, D.: Tyflos - A 2D Vibration Array for Sensing Dynamic Changes and 3D space for Blinds' Navigation. In: Proceedings of the 5th IEEE Symposium on Bioinformatics and Bioengineering (BIBE 2005). IEEE, Los Alamitos (2005), 0-7695-2476-1/05 $20.00 © 2005
12. Zelek, J.: Seeing by touch (haptics) for wayfinding. International Congress Series, vol. 1282, pp. 1108– 1112 (2005), 0531-5131/ D 2005, Elsevier B.V. All rights reserved
13. Meers, S., Koren, W.: A Vision System for Providing 3D Perception of the Environment via Transcutaneous Electro-Neural Stimulation. In: Proceedings of the Eighth International Conference on Information Visualisation (IV 2004), pp. 1093–9547. IEEE, Los Alamitos (2005), 1093-9547/04 $ 20.00
14. Hadjileontiadis, L.: - SmartEyes: an enhanced orientation and navigation system for blind or visually impaired people. Computer Society International Design Competition, Final Report (April 2003)
15. Nagarajan, R., Sainarayanan, G., Yacoob, S., Porle, R.: An Improved Object Identification For NAVI. Artificial Intelligence Research Group, School of Engineering and Information Technology, University Malaysia Sabah (2004)
16. Huang, N., Shen, Z., Long, S., Wu, M., Shih, H., Zheng, Q., Yen, N., Tung, C., Liu, H.: The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of The Royal Society Of London series A-Mathematical Physical and Engineering Sciences 454 (1971), 903–995 (1998)
17. Dafner, R., Cohen-Or, D., Matias, Y.: Context-based Space Filling Curves. In: Gross, M., Hopgood, F.R.A. (eds.) EUROGRAPHICS 2000, vol. 19(3) (2000)
18. Peano, G.: Sur une courbe qui remptit toute une aire plane. Math. Ann 36, 157–160 (1890)
19. Hilbert, D.: Aber die stetige Ubbildung einer Linie aufdie Flachenstuck. Math. Ann. 38 (1891)
20. Sezgin, M., Sankur, B.: Survey over Image Thresholding Techniques and Quantitative Performance Evaluation. Journal of Electronic imaging 13(1), 146–165 (2004)

Interaction in Mobility: The Evaluation of Interactive Systems Used by Travellers in Transportation Contexts

Christophe Kolski[1], Guillaume Uster[2], Jean-Marc Robert[3], Kathia Oliveira[1], and Bertrand David[4]

[1] LAMIH, Univ. Lille Nord de France, Univ.Valenciennes & Hainaut-Cambrésis, Le Mont-Houy, F-59313 Valenciennes Cedex 9, France
{firstname.name}@univ-valenciennes.fr
[2] Univ. Lille Nord de France, IFSTTAR, 20 rue Elisée Reclus, BP317, F-59666 Villeneuve d'Ascq Cedex, France
guillaume.uster@ifsttar.fr
[3]Ecole Polytechnique de Montréal, C.P. 6079, succ. Centre-ville, Montréal, Québec H3C 3A7
jean-marc.robert@polymtl.ca
[4]LIESP, Ecole Centrale de Lyon
36 avenue Guy de Collongue, F-69134 Ecully Cedex, France
Bertrand.David@ec-lyon.fr

Abstract. Transportation is a rich and complex domain for studying the use of interactive systems because of the diversity of travellers, activities, systems, and contexts of use, and the challenges that mobility represents for evaluation. In this paper we examine some new perspectives on transport and mobility and their impacts on evaluation. We propose to go beyond the evaluation of system utility, compatibility, accessibility and usability, and adopt the inclusive concept of User Experience. Finally we propose the use of a model-based engineering approach to take into account the variety of interactive systems, computing platforms, and media used in transport.

Keywords: Human-computer interaction, Mobility, Evaluation, transportation.

1 Introduction

This paper is about the evaluation of interactive systems used by travellers in transportation contexts. These contexts are heterogeneous since they include all means of transport (airplanes, ferries, trains, buses, subways, cars, bicycles, walk) and their respective home bases (airports, river ports, train/bus/subway/bicycle stations). What is striking here is the diversity of travellers, activities, interactive systems, and contexts of interaction, the diversity of situations that result from the combination of these elements, and the challenges that travellers' mobility represents for evaluation. This makes transportation a rich and complex domain for studying interaction [1].

The paper is structured as follows. We present some new perspectives on transport and mobility. Then we describe the sources of diversity in transport and we propose to

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 301–310, 2011.

evaluate the User Experience (UX) of travellers. We examine two activities which often occur in transport: learning and playing. Finally we describe some new perspectives offered by a model-based engineering approach.

2 Some New Perspectives on Transport and Mobility

Over the last years, there has been a diversification of forms of travels. Commuting is not anymore the most frequent form of travel, it has been replaced by occasional journeys. The planning of these journeys requires a lot of information and often comprises a part of uncertainty. For these journeys, private cars are still the principal mean of transport. Augmenting public transport and other green alternatives to private vehicles is the objective of sustainable mobility. But a major weakness of public transport is the lack of information. Indeed, several inquiries about occasional users of public transport mentioned this problem, for example, regarding bus routes and timetables, and the perception that the bus system is difficult to use and that information is difficult to access [2]. Furthermore, interchange of means of transport is perceived negatively by users and has also been identified as an area of public transport that should be improved [3]. Traveller information is a complex and vast research area; ATIS (Advanced Traveller Information Systems) is the common acronym to designate information systems in the area [4]. There are two main concerns in these systems: technology and user needs.

Technology can make the information accessible through a large diversity of media. For instance, during their trip, travellers can use an onboard information system through embedded screens (for instance, in a bus, a train) as well as they can use their own terminals.

User needs can be investigated through different data collection methods, including panels of travellers. Panels allow researchers to use several methods such as direct interviews, phone interviews, and focus groups to collect data [5].

Results of studies on different travellers indicate that they adapt their behaviors to travel contexts at each step, and tend to focus their attention on only a few pieces of critical information at a time. They need the right information at the right time, and relevant information should be provided in advance, before it is time to decide. Therefore, it is necessary to provide a coherent and efficient spatio-temporal system to meet travellers' needs at all stages of the journey [5]. Ergonomic studies can provide design specifications for the user interface (UI) of such a system [6], and accurate information can be delivered by the personalization of user content [7].

A travel is not only an itinerary from point A to point B, it is also the moment for travelers to do different activities such as sleeping, reading, listening to music, playing games, sending text messages, making phone calls, working, talking to other passengers, gazing out of the window, watching people, etc. [8]. Furthermore, during their trips, travellers pass through several transit spots where there is a continuous flow of information on transport, commercials, services, etc.

The principal success factor for sustainable mobility is to provide good quality interactive services to citizens and customers. This can be achieved on the base of careful observations of travellers at different times and different locations, for example onboard a train or a bus (Fig. 1). A Living Lab offers a good environment for making such observations and testing new equipments and services. It is a user-centred open innovation ecosystem, often operating on a specific territory, which allows one to do research and innovation through exploration, experimentation, and evaluation of innovative ideas, services, and equipments, in real life use cases and with numerous travellers. Furthermore, it allows one to collect data on delays, traffic, destinations, locations of the vehicles, infotainments, local news, touristic or cultural information, etc.

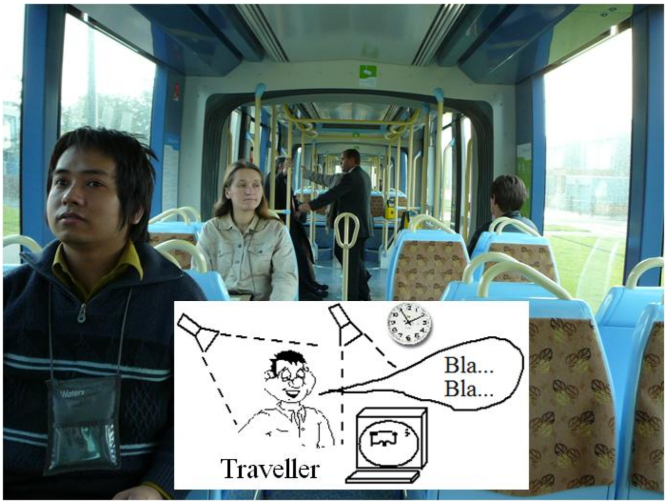


Fig. 1. Living Labs as new evaluation means in transportation contexts

3 Diversity of Usage Conditions in Transport and Evaluation of UX

Here we analyze the diversity of usage conditions in transportation contexts and its impacts on the evaluation of interactive systems, and we plead for the opportunity to evaluate the User Experience (UX) of travellers. Let’s start with diversity which comes from the interaction of four categories of elements:

- Diversity of travellers. This is remarkable, be it in an airplane, a ferry, a bus, a train, a metro, or in the street. Travellers reflect the entire population and include those with specific needs such as the handicapped, elderly, illiterate, colour-blind, people from different languages and cultures, with low technical abilities, with no or little knowledge about the system (because they use it for the first time or occasionally), with their own interactive systems that they expect to be able to use in conjunction with those of carriers. There are impacts on evaluation: in accordance with the objective of universal design, accessibility of systems should be guaranteed to almost all travellers (this is more and more imposed by regulations), usability of interactive systems should be tested with representatives of each category of travellers to make sure that these systems are easy to use, and systems compatibility should be ensured for the benefit of travellers who carry their own systems and expect to be able to use them in conjunction with those of different carriers

- Diversity of activities. This was mentioned above. Travellers will perform different activities when using systems of different carriers and their own systems. These activities could be related to the planning and organization of the trip itself, and to work or leisure done during transport or transit. They can be very diversified as shown in these examples: search for information (e.g., on costs, trajectories, arrivals and departures, connections), read dynamic displays, make reservations, register, buy tickets, make electronic payments, show a reservation, valid a ticket, get, renew or recharge a payment card, read and send emails, get a receipt, write texts, play with video games, listen to music, etc. Some activities are connected to others when they have prerequisites and successors, and will be perceived as such by the travellers. Similar activities, with minor differences, will be done on different interactive systems and in different contexts: they can be called generic and correspond to activity patterns. These observations have impacts on evaluation: adequate functionalities should be available in systems of carriers to perform these activities, the procedures imposed by these systems should be short, easy to follow, and consistent, similar (or generic) activities done on different systems and in different contexts should have similar procedures, and UIs that support these functionalities and procedures should be transparent, consistent, and provide adequate guidance to travellers.
- Diversity of systems. Travellers will use systems of different carriers as well as their own interactive systems to perform their activities. Here are several examples of such systems, some being interactive and others not: ticket dispenser, information terminal, registration terminal, dynamic visual information displays, automatic teller to refill a payment card or pay parking fees, bank teller, personal computer, smart phone, etc. These systems could be used by the same traveller in a short period of time and in different locations; their rate of use will differ widely depending on the category of travellers (occasional, average, frequent). These observations have impacts on evaluation: these systems should be evaluated as parts of a global technological ecosystem, i.e. as if they were in relation with each other. To compensate the lack of standardization, their external consistency should be emphasized in order to reduce travellers' cognitive load, minimize learning, prevent errors, and improve human performance and satisfaction. This means that the task scenarios behind the usability tests should include different activities, interrelated or not, done successively by a same traveller on different systems.
- Diversity of contexts. Travellers will be using systems in very different contexts. They could be in a city or a country where the language and culture are different from theirs and represent a barrier for the interaction with systems; when using these systems, they could be sitting, standing, or walking; in a hurry because of an imminent departure, or under pressure because of people waiting behind them; in a state of great fatigue because it is late or after a long trip; in a crowded and noisy place; they can be performing an activity for which security is important (e.g., payment); they could be using a system with one or two hands, and be manipulating objects during the interaction (e.g., a travel card, a credit card) while watching their suitcases; they could be alone, in couple, or in family; they could

be using one of their systems in conjunction with those of carriers (e.g., show the reservation code on their mobile phone at the booking desk). This reality has impact on evaluation: interactive systems should be tested in the field, not only in a laboratory, with real travellers doing real activities in conditions that represent the diversity and richness of use contexts.

Given the diversity of conditions that can be found in transportation contexts and the challenge it represents for the design of good interactive systems, it is important to adopt a user-centered design approach [10] and thus put emphasis on usability of interactive systems. This will help designers to realize systems that are more effective and efficient, easier to learn and use, and more satisfying for the users; these advantages are at the core of the definition of usability [11]. Despite the necessity of usability (and other qualities such as utility, compatibility, accessibility) of interactive systems in transportation contexts, we argue that it is not sufficient. It is an instrumental quality in a system that helps users to perform tasks and achieve goals. To a large extent it is simply a hygiene factor, i.e. a factor that has to be taken into consideration in order not to create dissatisfaction among users, and not a motivational factor, which is much more positive and attractive for users.

In a context of competition, where carriers aim to attract new travellers and establish their loyalty through good pricing and high quality services, and where travellers spend much time in transport and are critical towards the products and services they use in this domain, it is important to put emphasis on the global satisfaction of travellers. As a consequence, we propose to go beyond usability with its focus of UIs, and adopt UX with its focus of users (or travellers). It is not in the scope of this paper to present UX at length, we will rather present a few highlights that show the relevance of this concept in transport.

ISO 9241-210 (20101, clause 2.15) defines UX as: « A person's perceptions and responses that result from the use and/or anticipated use of a product, system or service ». Another definition is given by [12]: « UX is a multidimensional construct that defines the overall effect over time on the user of interacting with a system or service in a specific context ». The main features of the UX are the following: multidimensional, holistic, subjective, connected to the overall impression on the user, resulting from his/her interactions with the product and from a combination of anticipation and actual experience, dynamic (it changes over time at each point of user's contact with the product or service), situated in context (depending on location, time, other human interactions, opportunities and constraints, incidents, issues), and it can be studied at different levels of granularity [13,14]. UX is global and takes into account non-instrumental qualities of system, such as aesthetics, sense-making, the capacity to generate pleasure and emotions that were not considered in usability. It has been rapidly adopted by the communities of design, HCI and human factors, at the beginning of the years 2000. It is the object of an abundant literature and it represents some major challenges for researchers, namely understand the process and evolution of UX in time, and develop methods and tools for the evaluation of UX [15]. The global approach of UX seems essential for a more complete evaluation of interactive systems in transportation domain.

4 Learning and Playing in Mobility

In this section we examine two activities which occur often in transportation contexts: learning and playing. They appear in different terms and involve different people (Table 1). They can occur during travels done individually or in group by car, bus, train, plane, boat, etc. Two categories of people are involved: active or passive actors. Active actors are drivers or crews who will be interested in training as it is related to their work and allows them to improve their knowledge of the task (e.g., rules of operations, procedures about incidents or accidents) and of the equipments they control. Here we talk either of vocational training (for learning techniques or skills), just-in-time learning (for learning things when we just need them for performing an activity), contextual learning (related to a particular space, situation, period of time), or a combination of these. Passive actors are the passengers; let's consider two groups: adults and teenagers. All of them are looking for the best way to spend the transportation time, be it through entertainment (e.g., watch a video, play a video game, listen to music) or some structured learning activities, especially in a context of learning all-along the life. Both groups could be involved in vocational training or contextual learning. On the other hand, only teenagers (and young schoolchildren) will be associated to "anti uproar activities" (to be busy during a long trip in group) and "to do or to finish homework".

Table 1. Learning and playing activities

		Passengers	
Learning and playing activities	**Carriers**	**Adults**	**Teenagers**
Use appropriately transportation time		X	X
Anti uproar activity			X
Entertainment		X	X
Vocational training	X	X	X
Just in time learning	X		
Contextual learning	X	X	X
Contextual & just in time learning	X		
Do or finish homework			X

Figure 2 shows several examples of interactive devices embedded in planes, cars and buses, and of devices carried out by travellers (e.g., laptop, tablet, Smartphone, DVD reader, portable gamer, view goggle). They support individual or collaborative learning and playing activities through a connection to a net. They are often located on the back of a seat of a plane, a car or a bus, and they use a tactile screen or command panel. They use a WIMP or post WIMP interface with gestural, vocal, or multimodal interactions. New forms of interactions provided on new mobile or wearable devices, such as the portable Kinect and WII joysticks, started to appear and should be available in the near future. It is important to evaluate their efficiency and usability, the degree of comfort for users, and their impact on neighbors in transportation context.

Fig. 2. Learning and playing devices

5 New Perspectives Offered by a Model-Based Design Approach

In this section we propose the use of a model-based design approach in transport because of the variety of interactive systems, computing platforms, and medias (e.g., audio, video, image, text) in use. Here is an example. Let's consider the scenario of a person planning a trip by train at home with a PC. Then, while going to the train station s/he gets stuck in a traffic jam. S/he checks train schedules to find another departure, using a PDA, and changes the reservation. When arriving at the train station, s/he goes to the kiosk and uses a terminal to get the ticket. Here, three kinds of platforms were used and different contexts of use have to be taken into account. They represent new challenges for the design of interactive systems: produce systems that can be customized or adapted to all types of platforms, and respect the characteristics and location of the user when interacting with the system.

Model-Driven Architecture (MDA) [23] has gained importance as a paradigm that advocates the design of systems at a high level of abstraction. In MDA, models play a more direct role in software production, being amenable to manipulation and transformation by machine with the goal of final code generation. In general, this is done by the definition of three types of models corresponding to three abstraction levels: the computation independent model (CIM) which focuses on the system requirements; the platform independent model (PIM) which specifies the degree of platform independence that is appropriate with different platforms; and, the platform specific model (PSM) which combines the specifications in the PIM with the details that specify how that system uses a particular type of platform. Transformations are used to convert a model into another model (from CIM to PIM, and from PIM to

PSM) of the same system, up to the code generation. MDA has also being applied to UI design (see examples in [22, 24, 25]). UI design is done independently of the platform and then implemented. Models in the three levels and transformations of models should be done considering the context of use of the system, including the user, the platform and the environment [17]. Even though MDA seems to be an appropriate technology for the design of interactive systems in the transportation domain, two important questions come out: (1) How to consider the particularities of transportation systems with this paradigm? (2) How to evaluate the quality of the user interaction given the high level design model?

The first question is related to the worry of making the transport system usable by everyone and in different contexts. So, for instance, if the traveller is disabled, the system should propose direct itineraries with short walking distances for connections. It should also propose seats that consider his/her disability. In the same vein, if the user informs the system that s/he would like to do personal activities (such as pay bills, do shopping), the system should propose itineraries where services are offered at the connection points. To develop systems that consider all this particularities, we should include explicit models to capture domain knowledge and context information (about use, platform and environment). How to define these models? Which information to consider? Some propositions of context modeling can be found in the literature [20, 21]. The use of domain ontologies to organize the knowledge about transportation system has also been proposed [18]. Once these models are included in the design phase, it is essential to perform quality assessments to make sure that the particularities of this domain were really taken into account.

For the second question: since in MDA we talk about models and independence (at some level) or particularities of platform, we consider that the quality models can be generic at some point and should be specified when getting to the final UI (PSM level and final code). Besides, quality is considered multidimensional and depends on the product to evaluate and the perspective of evaluation. Therefore, we should think about what to evaluate for each model and by whom (end users, designers, experts, etc.). For example since the transportation system should be developed to be used by everybody, we should not only focus on *usability* but on *accessibility*. Usually, quality models are organized in a hierarchy view (tree view) of quality characteristics, from more general to more specific ones (for example, standards like ISO 9126 [19] have characteristics, sub-characteristics and metrics). The problem is when we go to more details in this hierarchy, it is difficult to keep the generality but it is richer about what we should really evaluate. Therefore we should consider top characteristics that can be evaluated for all MDA levels (e.g., usability) and that can be refined in sub-characteristics/attributes appropriate to each model. For example, in the PSM level specific metrics should be defined for the platform or a family of platforms that use one or several interaction modes (such as only speech or graphical and text integrated). Finally, we should analyze which methods to use for the evaluation of each quality characteristic of a model. This is a well-known problem in HCI, and with the MDA paradigm we should also think that we will evaluate generic models that will be used to produce systems for different platforms using different interaction models (speech system or basic textual system). This generality aspect must therefore be taken into account while defining evaluation procedures.

6 Conclusion and Perspectives

The evaluation of interactive systems in transportation contexts represents a great challenge for researchers and practitioners due to the diversity and complexity of aspects to be taken into account, and the rapid increase of technology. Indeed, this evaluation should cover technical factors such as systems capacity, compatibility, rapidity, security, reliability, robustness, integration, and technological modernity as well as human factors such as utility, accessibility, usability that have a direct impact on travellers' performance and satisfaction. Furthermore, we proposed that the evaluation go beyond these classical factors and cover the more global and inclusive travellers' User Experience.

We propose some ideas for supporting and boosting the evaluation of interactive systems in transport : adopt a global and integrated approach in order to evaluate each system as a component of a larger ecological system; promote the use of facilities such as the Living Lab described above in order to test rapidly and continuously new systems and equipments with real travellers in an environment very close to reality; and finally use the techniques of Web analytics (or equivalent when not on the Web) to collect data on the use of interactive systems, make diagnoses, and continuously improve the user interfaces.

Acknowledgements. This research work has been supported by CISIT, the Nord-Pas-de-Calais Region, and the European Community (FEDER).

References

1. Kolski, C. (ed.): Interaction homme-machine dans les transports - information voyageur, personnalisation et assistance. Hermes-Lavoisier, Paris (2010) ISBN 978-2-7462-3010-1
2. Beirao, G., Cabral, S.: Understanding attitudes towards public transport and private car. A qualitative study, Transport Policy 14, 478–489 (2007)
3. Hine, J., Scott, J.: Seamless, accessible travel: users' views of the public transport journey and interchange. Transport Policy 7, 217–226 (2000)
4. Lyons, G., Avineri, E., Farag, S., Harman, R.: Strategic review of Travel Information Research, final report to the department for transport, contract TDT/149 (2007)
5. Patesson, R., Lecomte, N.: User needs analysis: an analysis based on the study of activities of the traveller. In: Proc. 8th Int. Conference on Human-Computer Interaction, Munich (1999)
6. Mouloudi, A., Lemarchand, C., Valentin, A., Morizet-Mahoudeaux, P.: User information guidance in public transportation systems. In: Proc. 11th International Conference on Human Computer Interaction, Lawrence Erlbaum Associates Inc., Las Vegas (2005)
7. Brossard, A., Abed, M., Kolski, C., Uster, G.: User modelling: the consideration of the experience of time during journeys in public transportation. In: Proc. ACM Mobility Conference 2009. Nice, France, ACM Press, New York (2009)
8. Lyons, G., Jain, J., Holley, D.: The use of travel time by rail passengers in Great Britain. Transportation Research A: Policy and Practice 41(1), 107–120 (2007)
9. Uster, G., Juguet, S., Talon, G.: The Viatic concept information technology for intelligent travellers. In: Proc. 9th IEEE Int. Conference on ITS Telecommunications, pp. 569–573 (2009)

10. Gould, J.D., Boies, S.J., Ukelson, J.: How to design usable systems. In: Helander, M., Landaeur, T. (eds.) Handbook of HCI, pp. 231–254. North-Holland, Amsterdam (1997)
11. ISO DIS 9241-210:2010. Ergonomics of human system interaction - Part 210: Human-centred design for interactive systems (formerly known as 13407). International Standardization Organization (ISO), Switzerland (2010)
12. Robert, J.-M., Lesage, A.: From usability to user experience with user interfaces. In: Boy, G.A. (ed.) Handbook of Human-Machine Interaction, Ashgate, U.K (to appear, 2011)
13. Robert, J.-M., Lesage, A.: Designing and evaluating user experience. In: Boy, G.A. (ed.) Handbook of Human-Machine Interaction, Ashgate, U.K (to appear, 2011)
14. Law, E.L.C., Van Schaik, P.: Modeling user experience, An agenda for research and practice. Interacting with Computers 22, 313–322 (2010)
15. Vermeeren, A.P.O.S., Law, L.-C., Roto, V., Obrist, M., Hoonhout, J., Väänänen-Vainio-Mattila, K.: User Experience Evaluation Methods: Current State and Development Needs. In: Proceedings NordiCHI 2010, October 16-20, pp. 521–530 (2010)
16. Vink, P., De Looze, M.P.: Crucial elements of designing for comfort. In: Schifferstein, H.N.J., Hekkert, P. (eds.) Product Experience. Elsevier, CA (2007)
17. Calvary, G., Coutaz, J., Thevenin, D., Limbourg, Q., Bouillon, L., Vanderdonckt, J.A.: Unifying Reference Framework for Multi-Target User Interfaces. Interacting with Computers 15(3), 289–308 (2003)
18. Houda, M., Khemaja, M., Oliveira, K., Abed, M.: A public transportation ontology to support user travel planning. In: IEEE Proc. of Research Challenges in Inf. Science (2009)
19. ISO/IEC 9126-1: Software engineering - Product quality - Part 1: Quality model (2001)
20. Kim, E., Choi, J.: An Ontology-Based Context Model in a Smart Home. Computational Science and Its Applications, 11–20 (2006)
21. Lin, X., Li, S., Xu, J., Shi, W., Gao, Q.: An Efficient Context Modeling and Reasoning System in Pervasive Environment: Using Absolute and Relative Context Filtering Technology. In: Fan, W., Wu, Z., Yang, J. (eds.) WAIM 2005. LNCS, vol. 3739, pp. 357–367. Springer, Heidelberg (2005)
22. OlivaNova, `http://www.care-t.com/products/index.html` (last access: 07/01/2011)
23. OMG: MDA Guide Version 1.0.1 (2003)
24. Sottet, J.-S., Calvary, G., Coutaz, J., Favre, J.-M.: A Model-Driven Engineering Approach for the Usability of Plastic User Interfaces. In: Gulliksen, J., Harning, M.B., van der Veer, G.C., Wesson, J. (eds.) EIS 2007. LNCS, vol. 4940, pp. 140–157. Springer, Heidelberg (2008)
25. Vanderdonckt, J.: A MDA-Compliant Environment for Developing User Interfaces of Information Systems. In: Pastor, Ó., Falcão e Cunha, J. (eds.) CAiSE 2005. LNCS, vol. 3520, pp. 16–31. Springer, Heidelberg (2005)

Evaluation of Wayfinding Performance and Workload on Electronic Map Interface

Ya-Li Lin[*] and Cheng-Han Wang

Department of Statistics, Tunghai University,
Taichung, Taiwan 40704
yllin@thu.edu.tw

Abstract. Electronic map (E-map) is important to support wayfinding, especially for finding unfamiliar routes in the current Web mapping service applications. This study examined the effects of four designing factors including Map Size, Map Type, Direction key, and Zoom function as well as the ability of sense of direction (SOD) and gender differences on the wayfinding performance for a simulated E-map interface. The results indicate that participants with a better SOD would have the faster response time in average and would lower overall workload for target task. Furthermore, participants would have higher workload as the response time is increasing. The interaction effect of SOD and map type would affect the mean response time for target and direction tasks. Participants with good SOD using mixed map have the faster mean response time than ones of poor SOD. For direction task, males with good SOD and good SOD using E-map would have faster mean response time. In addition, both males using mixed map and females using E-map would have faster mean response time.

Keywords: Electronic map (E-map), Wayfinding Performance, sense of direction, NASA-TLX Task Load Index.

1 Introduction

With the advent of advanced spatial information technologies, mobile devices equipped with E-map (electronic map) service applications and/or GPS (Global Positioning System) receivers are common in the daily lives for mobile users. Wayfinding is a cognitive psychological process for finding pathway from an origin to a specified destination [13]. Paper maps have traditionally played major roles in convening spatial information and guiding people around in space. Literature on map learning has shown that using map is not an easy task for children and even for adults [10]. With respect to navigational aids, the study of Ishikawa, et al. [7] examined the effectiveness of GPS-based mobile navigation system in comparison to paper maps and direct experience of routes. Their results showed that GPS users traveled longer distances and made more stops during the walk than map users and direct-experienced

[*] Corresponding author.

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 311–320, 2011.

participants. In addition, various presentation formats of spatial information have been developed, including verbal navigational directions, static maps, interactive maps, 3-D visualizations, animations, and virtual environments [12]. Dillemuth [1] yielded faster travel speed and fewer navigation errors with a generalized map than with an aerial photograph. Some people can readily find their way back to a starting point along a route they have only experienced once, whilst others can only do this with considerable difficulty. Environmental spatial abilities have shown large individual differences [4, 8]. The individual difference between people is so called SOD (sense of direction). The results based on the study of Kato and Takeuchi [9] revealed that good SOD participants (GSD) showed much better performance on route learning than poor SOD participants (PSD). In addition, GSD made more flexible use of effective strategies than PSD. In addition, self-report measures of SOD have been found to predict objective measures of these abilities quite highly [5].

Google Maps and Yahoo! Maps are currently two popular Web mapping service applications. They offer street maps, a route planner for traveling by foot, car, or public transport and an urban business locator for numerous countries around the world. The usability of Google Map and Yahoo! Map interfaces has been evaluated using cross over design and the after-experiment questionnaire for user interface satisfaction (QUIS) [11]. The results of their study indicated Google's E-Map was rated as the highest satisfaction; however, paper map was rated as the lowest satisfaction. There are significant associations between gender and icon recognition. Male participants showed the higher percentage of correct recognition than females.

The functionality of E-map has been greatly advanced by the current technology of interactivity. It can be zoomed in and out, rotated without affecting the ratio of display, as well as easily combined with satellite images, aerial photographs and other sources of information to improve the user's understanding of the geographic database. In this study, the effects of four designing factors on the visual performance for three wayfinding tasks would be examined to provide an optimal user-centered interface for E-map. In addition, subjective assessments of SOD and NASA-TLX task load would be implemented by self-rating questionnaire after experiment.

This study differs from previous ones in that it concentrates on the research concerning designing factors affecting the wayfinding performance. To evaluate the usability of E-map, a simulated E-map interface is designed to collect the response time to complete the tasks of finding a target, identifying qualitative direction, and recognizing qualitative distance. In addition, the correlation between SOD, wayfinding performance, and overall workload would be investigated. The results of quantitative measurements and subjective assessments will be used as the guidelines to provide an ergonomic design and to meet the demands of usability for Web mapping service applications.

2 Method

2.1 Participants

Sixteen undergraduate and graduate students (8 females and 8 males) voluntarily participated in the experiment. Their ages ranged from 21 to 26 years old, with a mean of 24 years and standard deviation of 1.6 years. None of the participants had experience using E-map before the experiment. They all had normal vision or corrected vision reach at least 0.8 and no color-blindness. The requirement to be a participant is to leave alone PC before the formal experiment.

2.2 Materials

Apparatus

There is a 17-inches TFT-LCD monitor (1280×1024 pixels) and Intel(R) Core(TM) 2 Quad CPU Q6600 desktop computer (CPU 2.40 GHz, 1G RAM) with a headphone and a microphone in the laboratory. Optec 2000 Stereo Optical Vision Tester is used to measure vision acuity and examine the color blindness. TAKEI Digital Flicker TK502 is used to measure critical fusion frequency before and after experiment. Macromedia Flash 8 is used to design the simulated E-map interface. In addition, CyberLink Stream Author 3.0 is used to record the process of operating the system during the experiment. A digital video camera recorder (SONY DCR-PC330) is used to record the overall process of experiments and after-experiment questionnaire. In addition, the luminance of experimental lab is 487~611 lux measured by Lutron LX-101 Lux meter.

Design of Simulated E-Map Interface

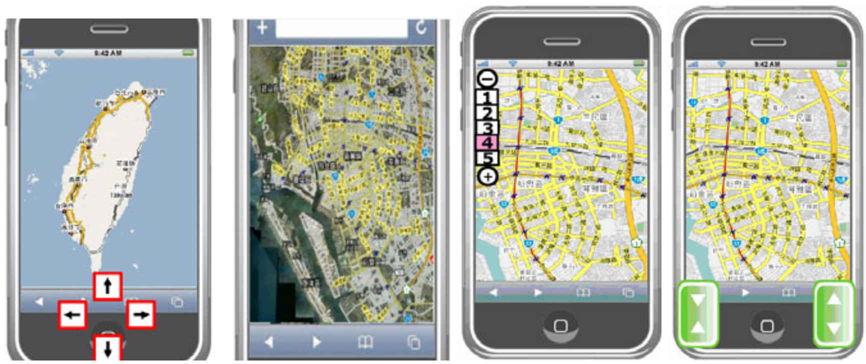
A simulated E-map interface is designed in this experiment. The sources of map images are cited from Google maps [2]. The study area is in the surrounding area of Kaohsiung City, Taiwan. Participant's current position is fixed on the same location as the starting point. The mapped area was dynamically updated as the user moved in space. The starting point and the goals were not always shown together on the map, that is, participant has to move down (head to southern) to the goal location and then use the zooming icons to find the target location. The location of the goal became visible as the user moved on.

The design factors used in the simulated E-map interface including two sizes of map interface (factor "Size", large size for desktop--468x326 pixels vs. small size for mobile device--231x331 pixels), two types of map (factor "Type", E-Map vs. Mixed map,), direction keys (factor "Key", Yes/No,), and hierarchical zoom icons (factor "Zoom", Yes/No). Mixed map means an E-map plus satellite map. The illustration of function of direction key is shown in Figure 1 (a) and 1(b). Figure 1 (a) is an E-map with the traditional up-down direction key and Figure 1 (b) is a mixed map with finger-touch design which means without using the movement of direction key. The illustration of function of hierarchical zoom key is shown in Figure 1(c) and 1(d). Figure 1 (c) is an E-map with traditional zoom icon using +/- and the corresponding numbers and Figure 1 (d) is an E-map with figure-touch design which means without using the clicking icons of zooming in and out.

Sense of Direction Scale

Based on Santa Barbara Sense of Direction Scale (SBSOD) [5], 10 questions of spatial and navigation using Likert's seven-point scale were filled out before the formal experiment. Participants rated each question by circling a number ranging from 1 (strongly disagree) to 7 (strongly agree). Seven out of ten questions are stated positively, e.g., "My sense of direction is very good," "I am very good at reading maps." The other three questions are stated negatively, e.g., "I have a poor memory for where I left things," "I very easily get lost in a new city." The answers will be reversed to positive statement so that a higher score means a better SOD. The relationship of

SOD and response time and the corresponding overall workload would be investigated for three tasks.



(a) (b) (c) (d)

Fig. 1. Illustration of map used in the experiment: (a) E-map with direction key, (b) Mixed map without direction key, (c) E-map with zooming icon, and (d) E-map without zooming icon. Source of map data: Google Map, http:// maps.google.com.tw/.

NASA-TLX Task Load Index

National Aeronautics and Space Administration Task Load Index (NASA-TLX) [3] is used to assess the subjective overall workload for three tasks. A multi-dimensional rating scale is proposed in which information about the magnitude and sources of six workload-related factors are combined to derive a sensitive and reliable estimate of workload. A rating scale of ten-point for six subscales, which consist of mental demand, physical demand, temporal demand, performance, effort, and frustration level. The relative importance of the six component factors to each subject's personal definition of workload was determined in a pretest. All possible pairs (totals 15) of the six factors were presented in a different random order to each subject. The member of each pair selected as most relevant to workload was recorded and the number of times each factor was selected was computed. The resulting values could range from 0 (not relevant) to 5 (more important than any other factor). The more important a factor was considered to be, the more weight the ratings of that factor were given in computing an average weighted workload score (WWL) for each experimental condition. An average of these six subscales, weighted to reflect the contribution of each factor to the workload of a specific activity from the perspective of the rater, is proposed as an integrated measure of overall workload.

2.3 Design of Experiment

An orthogonal array experiment $L_{16}(2^{15})$-- similar to a four-factorial experiment with single replicate will be used to collect the wayfinding performance of visual search. Three tasks will be assigned to each participant, which are (1) to find the targets, (2) to identify the cardinal directions, and (3) to identify the approximate distances. Cardinal directions are based on 8-sectors model (North, East, South, West, North-East, South-East, South-West, and North-West), while approximate distances correspond to a set of ordered intervals that the order among symbolic distance values describes distances from the nearest to the furthest [6]. Time to correctly complete the target task, time to correctly identify the direction task, and time to correctly identify the distance task will be collected based on the simulated E-map interface. Designing factors included (1) Size (Large/Small), (2) Type (E-map/ Mixed map), (3) Key (Yes/No), and (4) Zoom (Yes/No).

2.4 Procedure

At the beginning of the experiment, naive participants were asked to follow the instructions to learn how to operate the simulated E-map interface. They have to practice using the device until they knew how to use it. Before the formal experiment, they were asked to fill out the sense-of-direction questionnaire. One out of sixteen treatment combinations was randomly assigned to one of the participants, and the experimental sessions of wayfinding tasks began. After one of three wayfinding tasks being done, NASA-TLX rating questionnaire was separately filled out by each participant.

3 Results

3.1 Sense of Direction

The self-rating of sense of direction (SOD) is calculated by summing up 10 SOD questions as SOD score. The descriptive statistics are shown in the first row of Table 1. Although the mean SOD of males (43.25) is higher than ones of females (35.1), it is lacking of sufficient evidence to support the gender difference in sense of direction.

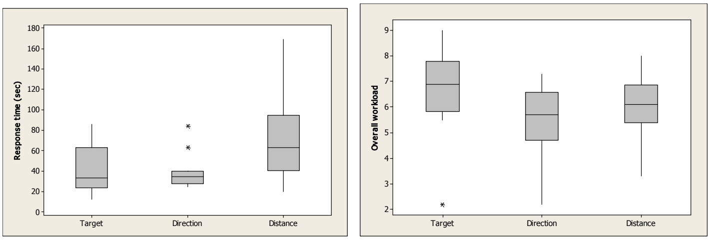
Table 1. Descriptive statistics of SOD, response time (RT), and overall workload (OW) for three wayfinding tasks (n=16)

Variables	Mean	Std Dev	Min	Q_1	Median	Q_3	Max
SOD	39.2	10.4	19.0	34.0	36.5	47.0	56.0
RT of Target[1]	41.1	23.0	12.8	23.9	33.6	63.2	86.0
RT of Direction[1]	38.1	15.3	25.0	27.8	35.0	39.8	84.0
RT of Distance[1]	70.8	38.3	20.0	40.5	63.0	94.5	169.0
OW of Target	6.7	1.6	2.2	5.8	6.9	7.8	9.0
OW of Direction	5.5	1.4	2.2	4.7	5.7	6.6	7.3
OW of Distance	6.0	1.2	3.3	5.4	6.1	6.9	8.0

Note 1: the unit of cell is second.

3.2 Wayfinding Performance

Wayfinding performance consists of: (1) response time of correctly finding the targets, (2) response time of correctly identifying cardinal directions, and (3) response time of correctly identifying the approximate distances. The descriptive statistics are shown in Table 1. Mean response times of three wayfinding tasks are significantly different based on two-way ANOVA blocked by participants (F=8.38, p=0.001). Post hoc paired comparisons showed that the mean response time of identifying distance task was longer than that for finding the targets and that for identifying direction (the values shown in the second, third, and forth rows of Table 1 and illustrated in Fig. 2(a)).



(a) (b)

Fig. 2. Box plots of response time in (a) and NASA-TLX overall workload in (b) for three wayfinding tasks

Based on the self-report scale of sense of direction (SOD), the result of correlation between SOD and response time of three tasks is shown in Table 2. It indicates a moderately negative correlation (r=-0.62, p-value=0.011) between SOD and response time of finding targets, that is, the response time is decreasing as SOD is increasing. However, there are only weakly negative correlation between SOD score and response time of identifying direction as well as identifying distance.

Furthermore, regrouped SOD into two groups--GSD and PSD (named as SODG). The results of ANOVA (Table 3) indicate there is statistically significant interaction effect of SODG *Type (F=5.53, p=0.037). In Fig. 3, the combination of GSD using mixed map has the fastest mean response time (21.56 seconds) for target task.

Table 2. Pearson correlations (r) between response time (RT) and overall workloads for three wayfinding tasks

r	1. SOD	2	3	4	5	6
2. RT of Target	**-0.616***					
3. RT of Direction	-0.277	0.140				
4. RT of Distance	-0.268	0.429	-0.113			
5. OW of Target	**-0.538***	**0.657****	-0.050	0.280		
6. OW of Direction	0.232	-0.381	-0.218	0.170	0.236	
7. OW of Distance	0.275	-0.009	-0.259	0.190	0.480	**0.669****

*p<0.05;**P<0.01

Table 3. ANOVA table for target task

Source of variation	Sum of square	DF	F	*p*-value
SODG	2894.4	1	27.00	0.000**
Type	3127.7	1	29.18	0.000**
SODG*Type	592.8	1	5.53	0.037*
Error	1286.4	12		
Total	7901.4	15		

*$p<0.05$;**$P<0.01$

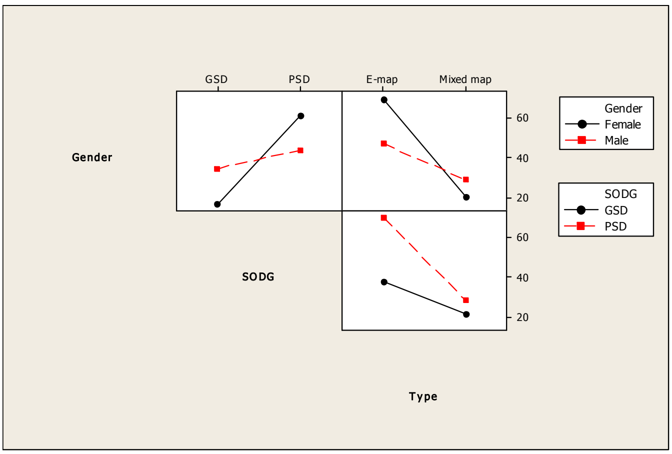


Fig. 3. Interaction plots for target task

Table 4. ANOVA table for direction task

Source of variation	Sum of square	DF	F	*p*-value
Gender	14.06	1	0.69	0.429
SODG	266.70	1	13.03	0.006*
Type	59.73	1	2.92	0.122
Gender*SODG	284.91	1	13.92	0.005**
Gender*Type	1172.12	1	57.28	0.000**
SODG*Type	1517.25	1	74.15	0.000**
Error	184.17	9		
Total	3498.94	15		

*$p<0.05$;**$P<0.01$

The results of ANOVA (Table 4) indicate there are statistically significant interaction effects of Gender*SODG (F=13.92, p=0.005), Gender*Type (F=57.28, p=0.000) as well as SODG *Type (F=74.15, p=0.000). In Fig. 4, male GSD in average has faster response time (30.4 seconds) and GSD using E-map in average has faster response time (28 seconds). In addition, both males using mixed map and females using

E-map in average have faster response time, which is 32.25 and 32.5 seconds, respectively. Participants with a better sense of direction tended to shorter response time in this study. It is similar to the finding of Dillemuth [1]. It would be interesting to see how the ability of sense of direction affects people's wayfinding performance. This is in line with Ishikawa et al. [7] for their map group that a significant relationship between participants' sense of direction and wayfinding performance.

3.3 NASA-TLX Rating

The self-report rating based on NASA-TLX revealed significantly higher overall workload for target task. The descriptive statistics are shown in Table 1. There is a significant difference among the overall workloads of three tasks based on two-way ANOVA blocked by participants (F=5.66, p=0.008). The task of finding the target has the highest overall workload shown in Fig. 2(b). The result of correlation between SOD and overall workload of three tasks is shown in Table 2. It indicates a moderately negative correlation (r=-0.54, p-value=0.032) between SOD and overall workload for target task, that is, the overall workload is decreasing as SOD is increasing. In addition, participants with a slower response time tended to higher overall workload (r=0.66, p-value=0.006) for target task. There is a moderately positive correlation (r=0.67, p-value=0.005) between overall workload of recognizing direction and recognizing distance tasks.

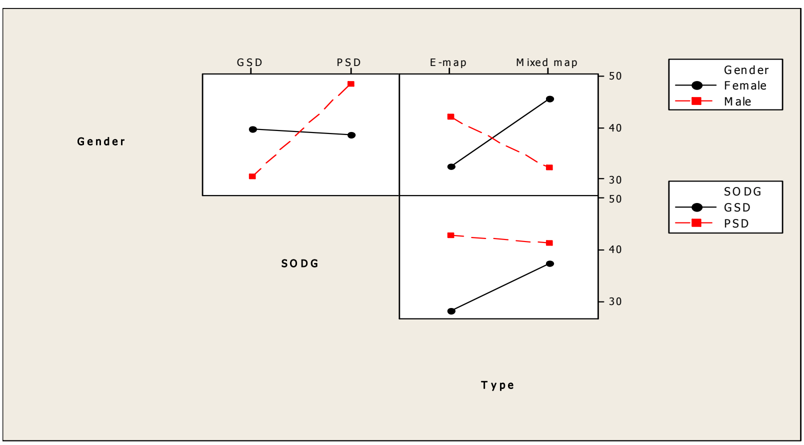
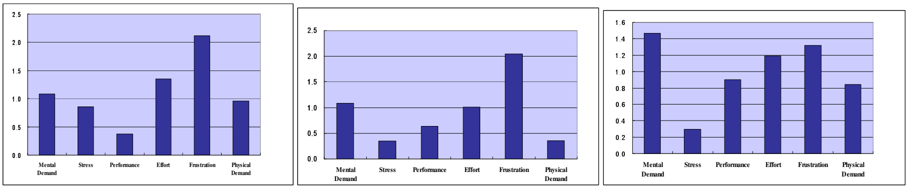


Fig. 4. Interaction plots for direction task

Six subscales including mental demand, physical demand, temporal demand, performance, effort, and frustration level were used to compute weighed task load index. The subscales of frustration, mental workload, and effort are regarded as the most important indices for tasks of finding targets and identifying direction based on the results of NASA-TLX rating scales in Fig. 5(a) and 5(b). However, the subscales of mental workload, frustration, and effort are regarded as the most important indices for task of recognizing distance based on the results of NASA-TLX rating scales in Figure 5(c).



(a) (b) (c)

Fig. 5. Average subjective rating of six workload related subscales for three wayfinding tasks: (a) finding target, (b) identifying direction, and (c) recognizing distance

4 Conclusions

In summary, participants with a better SOD would have the faster response time in average and would lower overall workload for target task. Furthermore, participants would have higher workload as the response time is increasing. The interaction effect of SOD and map type would affect the mean response time for target and direction tasks. Participants with good SOD using mixed map have the faster mean response time than ones of poor SOD. For direction task, males with good SOD and good SOD using E-map would have faster mean response time. In addition, both males using mixed map and females using E-map would have faster mean response time.

Acknowledgements. The support of the National Science Council, Taiwan (grant NSC 99-2221-E-029 -029) is gratefully acknowledged.

References

1. Dillemuth, J.: Map Design Evaluation for Mobile Display. Cartography and Geographic Information Science 32, 285–301 (2005)
2. Google Maps (2010), `http://maps.google.com.tw/`
3. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In: Hancock, P.A., Meshkati, N. (eds.) Human Mental Workload, pp. 239–250 (1998)
4. Hegarty, M., Montello, D.R., Richardson, A.E., Ishikawa, T., Lovelace, K.: Spatial Abilities at Different Scales: Individual Differences in Aptitude-Test Performance and Spatial-Layout Learning. Intelligence 34, 151–176 (2006)
5. Hegarty, M., Richardson, A.E., Montello, D.R., Lovelace, K., Subbiah, I.: Development of a Self-Report Measure of Environmental Spatial Ability. Intelligence 30, 425–447 (2002)
6. Hong, J.H., Egenhnfer, M.J., Frank, A.U.: On the Robustness of Qualitative Distance-and Direction Reasoning. In: Peuquet, D. (ed.) Proceedings of Autocarto, vol. 12 (1995)
7. Ishikawa, T., Fujiwara, H., Imai, O., Okabe, A.: Wayfinding with GPS-Based Mobile Navigation System: A Comparison with Maps and Direct Experience. Journal of Environmental Psychology 28, 74–82 (2008)
8. Ishikawa, T., Montello, D.R.: Spatial Knowledge Acquisition from Direct Experience in the Environment: Individual Differences in the Development of Metric Knowledge and the Integration of Separately Learned Places. Cognitive Psychology 52, 93–129 (2006)

9. Kato, Y., Takeuchi, Y.: Individual Differences in Wayfinding Strategies. Journal of Environmental Psychology 23, 171–188 (2003)
10. Liben, L.S., Kastens, K.A., Stevenson, L.M.: Real-World Knowledge Through Real-World Maps: A Developmental Guide for Navigating the Educational Terrain. Developmental Review 22, 267–322 (2002)
11. Lin, Y.L., Wang, C.H.: Evaluation of Electronic Map Interface on Web Pages for Naive Users. Journal of Ergonomic Study 12(1), 43–54 (2010)
12. Montello, D.R., Waller, D., Hegarty, M., Richardson, A.E.: Spatial Memory of Real Environments, Virtual Environments, and Maps. In: Allen, G.L. (ed.) Human Spatial Memory: Remembering Where, pp. 251–285. Erlbaum, Mahwah (2004)
13. Xia, J., Arrosmith, C., Jackson, M., Cartwrigh, W.: The Wayfinding Process Relationships Between Decision-Making and Landmark Utility. Tourism Management 29, 445–457 (2008)

Implementing Effective Tactile Symbology for Orientation and Navigation

Bruce Mortimer, Gary Zets, Greg Mort, and Curtis Shovan

Engineering Acoustics Inc., 406 Live Oaks Blvd., Casselberry, FL, USA
bmort@eaiinfo.com

Abstract. The sense of touch is an effective, but underutilized, human communication channel. In this paper we describe our research efforts towards optimizing a minimal tactile array for personal navigation and route guidance. There are several aspects to this problem. From an information transfer viewpoint, the question of tactor array size, dimension, location and display symbology requires careful consideration. Effective tactile display symbology involves providing information in an intuitive manner without adding to the cognitive loading of the user. Tactile information may be presented through spatial, temporal and signal variables. We have recently developed new wearable tactors that offer wide sensory capabilities to provide different "feeling" stimuli. These actuators are non-linear in that the salient characteristics for perception are linked to a complex drive stimulus. We have therefore developed a tactor activation design approach termed "TActions" (Tactile Actions) where patterns or sequence of individual tactile stimuli, each of which has its own characteristics and properties, are used to create tactile display symbology that a user can naturally associate with a particular function. These components provide display design frame work which we have used to demonstrate orientation and navigation.

Keywords: Tactile display, navigation, tactor.

1 Introduction

The sense of touch offers a relatively untapped and intuitive channel for communication and orientation. It is intrinsically linked with the neuro-motor channel, both at the reflex and higher cognitive regions, which make it naturally suited to localization tasks. Tactile arrays can be an effective communication modality even under situations where the conventional communication channels such as visual, audio and even the vestibular system become disorientated [1]. Further, according to multiple resource theory [2], parsing information across the input modalities can potentially alleviate sensory bottlenecks. Tactile displays offer additional advantages in that they are potentially covert, omnipresent and are omni-directional (the user does not need to be looking in a specific place to receive tactile feedback and it may be applied at almost any place on the body – in effect, act as "eye's in the back of the head").

Research [3] has also demonstrated that under appropriate conditions, tactile cueing yields significantly faster and more accurate performance than comparable visual or spatial auditory cues. This finding is relatively stable across a variety of

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 321–328, 2011.

body orientations, even when limited spatial translation is required and under physiological stress [4]. However, there are a number of critical human factor issues and hardware system limitations that must be considered in the design of practical tactile cueing systems. Our focus has been on optimizing wearable tactor actuator components [5] and tactile symbology for tactile cueing. In this paper we describe how new tactile hardware and systems constructs can be applied in personal navigation and route guidance.

2 Background to Navigation and Orientation

The informational requirements associated with navigation and orientation will depend on the environment, task and application. In complex environments, these requirements may require aspects of motion (velocity), position and orientation with respect to a frame of reference. It is very difficult to convey large amounts of information simultaneously and the tactile channel is particularly limited. Effective tactile displays should ideally provide navigation and orientation information intuitively without adding to the cognitive loading of the user. This usually requires careful bounding of the display (limiting the dimensional variables), designing specific tactile symbology and careful tactor array implementation.

Attentive observers are able to perform complex spatial translation and orientation mapping during navigation tasks, but this is usually less effective in dual task conditions or if the user is under stress (such as that encountered in many military applications). Therefore displays that are implicitly aligned with a world coordinate axis are preferred. The torso is usually aligned with the direction of intended motion and is also well suited to the presentation of multidimensional spatial concepts such as elevation and azimuth (with respect to body coordinates). This can be particularly useful in aviation, as demonstrated by the tactile situational awareness system (TSAS) [6] where 3-D orientation variables are needed. For navigation and orientation in many ground based tasks the plane of interest can be restricted to only the azimuth, greatly simplifying the array.

The simplest informational requirements for completing a navigation task are the direction to and distance from the waypoint, and this can be presented on a torso worn tactor array [7]. Directional information is naturally mapped to an appropriate sector on the torso and studies [8] have shown that an array of 8 tactors represents a reasonable compromise between resolution and accuracy. Typically sensor systems calculate and map the target waypoint location (with respect to the body axes) to the corresponding sector on the tactile belt. The user then turns to align to the waypoint, and the tactile cue moves to the front tactor to confirm the correct heading and orientation. Some information regarding the distance from the way point can be conveyed by modulating the pulse rate of the tactors (modulating faster when one is in close proximity to the target). Although tactile navigation has been shown to be useful [9], it is desirable to extend the capabilities of tactile displays beyond simple directional cueing and improve the overall effectiveness.

3 Tactile Array Displays

3.1 The Perception of Tactile Stimuli

There is a difference between feeling a stimulus, and being able to assign meaning to that sensation. The extension of tactile orientation cues to more advanced symbology is limited by the relatively narrow tactile channel bandwidth of the body [10]. Our model for the perception of tactile stimuli is shown in Figure 1.

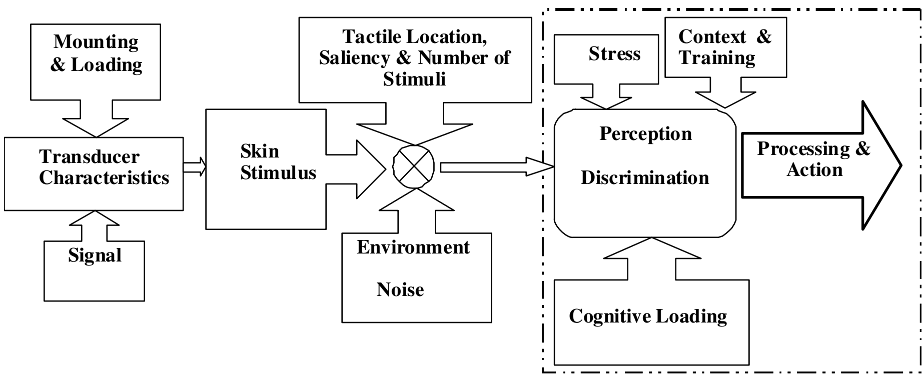


Fig. 1. Model for the perception of tactile stimuli – the transfer function of the transducer, environment, tactile stimuli, user workload, loading and task context all play a significant role in the tactile perception process

Actuators, especially light weight wearable transducers, are prone to loading and are seldom able to present wide-bandwidth vibratory signals when loaded against the skin. Therefore the first limitation in the tactile perception system is the vibrotactile actuator itself - the signal will be modified by the transfer function of the transducer. Perception of vibrotactile stimuli depend on body location, actuation area and the characteristics of the stimulus [11]. It is also known that attention influences sensory performance [12] and that distraction such as noise, masking from other vibratory sources and resource competition also potentially influence perception. Clearly ideal tactile display constructs should offer reliable and intuitively discernable cues within the intended application and environment.

3.2 Wearable Vibrotactile Transducers

Wearable vibrotactile transducers fall into two broad categories depending on their design and inertial reference [5]. The first is that of inertial shakers: A typical example would be the widely used eccentric mass motor or pager motor. The second category is that of a linear actuator which drives a contactor or moving element against the skin. In the inertial shaker, the rotation of the motor causes the housing to vibrate, stimulating the adjacent skin. However, this approach can be rendered ineffective through any additional mass loading, for example due to mounting.

Linear actuators such as Engineering Acoustics Inc's (EAI) C-2, C-3 and EMR tactors are shown in Figure 2. These tactors are configured such that when an electrical signal is applied, the "contactor" oscillates perpendicular to the skin, while the surrounding skin area is "shielded" with a passive housing. These linear actuator tactors provide a strong, point-like sensation that is easily felt and localized. In wearable systems, the overall mass of the actuator should be minimized. However, from a linear actuator transduction viewpoint, the mass of the housing is a "reaction mass" against which the contactor operates while displacing the skin load. Thus there is a careful tradeoff between contactor and housing dimensions and the actuator (housing and contactor) mass ratio's.

Fig. 2. The EAI C-2, C-3 and EMR vibrotactor transducers (left to right respectively)

Non-spatial tactile symbology includes modulating the operating frequency, amplitude, waveform, or temporal properties of a vibration signal. The skin comprises of a number of different types of mechanoreceptors that are sensitive to different frequency ranges and rates of adaptation [13]. Pacinian (PC) receptors have high vibration sensitivities and are perceived with a high degree of "urgency" and saliency. We have previously used mechanical resonances to improve vibrational performance in a linear motor design (C-2 and C-3 shown in Figure 2) optimized to the PC range (250 Hz). The EMR is a new motor based design with an operating frequency around 60-160 Hz. This design is able to produce a wide range of perceivable tactile features ranging from a strong "alert" (similar to the sensation of a C-3 or C-2), to a "soft" pressure pulse or "nudge". Control of the sensation parameters is via a complex mapping of the motor's drive voltage. Our explanation for this effect is that we believe that the vast majority of perceivable features (in short tactile pulses) are contained in the rise time, peak displacement amplitude, frequency and pulse envelope. Rapid turn-on pulse characteristics will result in a relatively broad Fourier spectrum, which will (depending on the pulse amplitudes and durations) excite PC type receptors as well as other mechanoreceptors, resulting in an "urgent" stimulus. Lower frequency stimuli produce sensations that are typically regarded as being less urgent provided the tactor turn-on characteristics are smooth or "gradual". Lower frequency vibration is part of the haptic environment and may therefore be regarded as being more "natural" but also more prone to being masked by environmental noise. Therefore we have designed the EMR to be capable of producing substantial peak displacements of up to 1.2 mm p-p (as measured against a phantom with the mechanical impedance of skin). In contrast, the C-2 (operating at 250 Hz) would typically only be driven to peak displacements of about 0.5 mm p-p owing to the relatively high PC channel displacement sensitivity.

3.3 Design Tools for Tactor Symbology

Structured tactile patterns can be used as messages where information is encoded into the signal by manipulating parameters of the vibration [14]. There are several signal parameters that can be modulated [15]. We have developed a tactor activation design abstraction approach and software tool termed "TActions" (tactile Actions). TActions are patterns or sequences of individual tactile stimuli (each of which has its own characteristics and properties). TActions can also have spatial relationships which can be easily assigned to reference a particular tactor in the array. TActions allow designers to easily create extremely complex patterns (with spatial and sequential pattern features).

Our software tool has two components; a design environment "TAction Writer" and a TAction management layer. The design environment provides a graphical environment for creating various tactor array geometries or "layouts" and an editing / configuration tool for designing multidimensional tactile patterns. The patterns can be created and saved as "TActions" which can also be reused. TActions can be assigned to application specific events or message contexts, providing a basis for developing common tactile symbology. Our overall design goal is that tactor display designers can use TAction library components and develop their own symbology that users can naturally associate with a particular function.

4 Orientation and Navigation Symbology

New capabilities in vibrotactile actuators and design tools facilitate more complex tactile symbology in orientation and navigation. For example, in navigation wayfinding tasks a TAction can be used to indicate that the waypoint target has been reached, or in other implementations indicate the potential for different route options, i.e. convey a predetermined context sensitive message. Tactile messages require some aspects of training and must be carefully designed with properties that are clearly distinct so that they can be reliably separated from other messages and directional cueing information. Changing the tactile salient features together with time varying spatial and rhythmic pattern implementations has been previously found to be intuitive in conveying military hand signals [16]. Thus this extends the capabilities of tactile displays beyond simple directional cueing to one of mission support, improving overall effectiveness.

It is instructive (especially for emerging hand held navigation applications) to also consider approaches for minimizing the tactile display array size. Generally orientation requires a tactile display with a complete spatial representation to be effective while navigation only requires the presentation of spatial directional course cues. Therefore tactile displays can be reduced in application specific implementations.

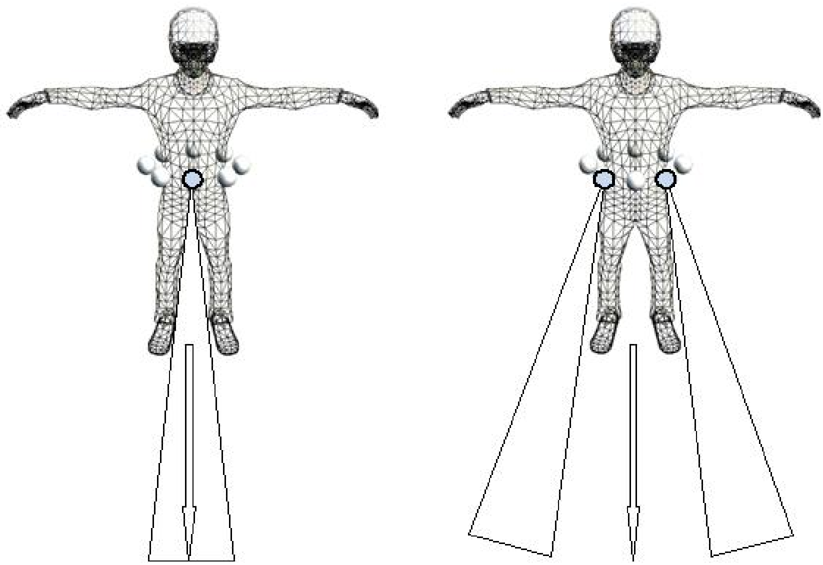


Fig. 3. Two distinct directional cueing for navigation modes that are possible with a torso worn tactile array; move to cue (left) and move away from cue (right). The arrow depicts the correct heading to the waypoint while the shaded circles represent an example of active tactors for each mode – as the user's orientation deviates from the heading either a new tactor sector is activated (move to cue), or a left or right error is activated (move away from cue).

Tactile spatial direction course cues can be either directional (move towards the cue) or, we can represent the information as an "error" (move away from the cue) as shown in Figure 3. In each case the tactor cue can be given proportional to the deviation from the course line (depending on the tactile symbology chosen). Both constructs have natural analogies; moving towards a cue is similar to tapping on a shoulder, while move away from a cue is similar to bumping into a real physical corridor barrier. Each approach has similar minimum array requirements but very different operating modalities; a directional course cue can be designed to be continuous i.e. the tactor pulse pattern is continuously firing, while a deviation cue is only presented if the course deviates beyond a threshold. The effectiveness of each implementation depends on training, multisensory display modalities and the confidence in the navigation system – continuous cues are beneficial in that there is a constant representation that the system is operating but there is also a risk for elements of change blindness [17].

Generally systems that are visually intensive (such as aviation) are better suited to navigation constructs that are error threshold based (move away from cue). This approach also allows multiple layers of information, such as threat or orientation, to be potentially represented on the tactile display (using the advanced tactor capabilities described previously). For example threats may be represented as a high priority, high frequency, high amplitude, pulse train stimulus. This is distinct from the navigation cues which may be lower priority, low frequency, slow rise time, high amplitude pulse train stimuli. Although these modes must be prioritized so as to prevent any simultaneous tactor mode displays, we successfully implemented this in TSAS (operators are able to easily act on the correct cue). Various solutions, including a hierarchy for the presentation of information such as threat, alerts, target, range, heading and navigation have been implemented using this approach.

Dismounted warfighter navigation applications are, in contrast, better suited to navigation constructs that are directional (move towards cue). With appropriate training, it is possible that a minimum of only three tactors are needed (center, left and right) for navigation – the navigation objective is to orientate and move such that the center tactor is activated. The salient characteristics ("urgency" and pulse repetition rate) can be modulated based on waypoint proximity. However, the restricted dimensions of the array and continuous navigation information will restrict the type and saliency of TAction message constructs.

5 Conclusions

Tactile displays present an intriguing and natural pathway for representing orientation and navigational information. However, there are several compromises in the system design and display interaction that must be carefully implemented (based on the specific requirements of a particular application). We have presented a simple model for understanding the effects of human factor and hardware on the perception of tactile cues. Our current research efforts are focused on optimizing tactor navigation systems that use each of the described tactile display modes. We have built a new range of tactile hardware and implemented software development tools that greatly increase the overall capability of system. In this work, we have illustrated how these capabilities can be used to design minimized tactile arrays for navigation. Future research will integrate multisensory information (especially visual) with tactile augmentation, and quantify the impact of different tactor array dimensions, and natural tactile symbology on task performance.

References

1. Raj, A.K., Suri, N., Braithwaite, M.G., Rupert, A.H.: The tactile situation awareness system in rotary wing aircraft: Flight test results. In: Proceedings of the RTA/HFM Symposium on Current Aeromedical Issues in Rotary Wing Operations, pp. 16-1–16.6. RTO NATO, France (1998)
2. Wickens, C.D.: Multiple resources and performance prediction. Theoretical Issues in Ergonomic Science 3(2), 159–177 (2000)
3. Elliott, L., Coovert, M.D., Redden, E.S.: Overview of Meta-analyses Investigating Vibrotactile Versus Visual Display Options. In: 13th International Conference on Human-Computer Interaction, San Diego, CA (2009)
4. Merlo, J., Stafford, S., Gilson, R., Hancock, P.A.: The Effects of Physiological Stress on Tactile Communication. In: Proceedings of the 50th Annual Meeting of the Human Factors and Ergonomics Society, San Francisco (2006)
5. Mortimer, B.J.P., Zets, G.A., Cholewiak, R.W.: Vibrotactile transduction and transducers. J. of the Acoustic Soc. of Am. 121(5), 2970–2977 (2007)

6. Rupert, A., Guedry, F., Reschke, M.: The use of a tactile interface to convey position and motion perceptions. AGARD Meeting on Virtual Interfaces: Research & Applications (1993)
7. van Erp, J.: Tactile navigation display. In: Brewster, S., Murray-Smith, R. (eds.) Haptic HCI 2000. LNCS, vol. 2058, p. 165. Springer, Heidelberg (2001)
8. Cholewiak, R., Brill, J., Schwab, A.: Vibrotactile localization on the abdomen: Effects of place and space. Perception & Psychophysics 66(6), 970–987 (2004)
9. Dobbins, T., Samways, S.: The Use of Tactile Navigation Cues in High-Speed Craft Operations. In: Proceedings of the RINA Conference on High Speed Craft: Technology and Operation, pp. 13–20. The Royal Institution of Naval Architects, London (2002)
10. Tan, H.Z., Durlach, N.I., Reed, C.M., Rabinowitz, W.M.: Information Transmission with a Multi-Finger Tactual Display. Perception and Psychophysics 61, 993–1008 (1999)
11. Cholewiak, R., Collins, A.: Sensory and physiological bases of touch. In: Heller, M.A., Schiff, W. (eds.) The Psychology of Touch, pp. 23–60. Erlbaum, Hillsdale (1991)
12. Eimer, M., Forster, B.: The spatial distribution of attentional selectivity in touch: evidence from somatosensory ERP components. Clinical Neurophysiology 114, 1298–1306 (2003)
13. Goodwin, A.W.: Paradoxes in tactile adaptation. Focus on vibratory adaptation in cutaneous mechanoreceptive afferents and time-course of vibratory adaptation and recovery in cutaneous mechanoreceptive afferents. J. Neurophysiol. 94(5), 2995–2996 (2005)
14. Brewster, S.A., Brown, L.M.: Non-visual information display using tactons. In: CHI 2004: Extended Abstracts on Human Factors in Computing Systems, pp. 787–788. ACM Press, Vienna (2004)
15. Jones, L.A., Sarter, N.B.: Tactile displays: Guidance for their design and application. Human Factors 50(1), 111 (2008)
16. Gilson, R.D., Redden, E.S., Elliot, L.R. (eds.): Remote Tactile Displays for Future Soldiers (Vol. ARL-SR-0152): Aberdeen Proving Ground, MD 21005-5425 (2007)
17. Ferris, T.K.: Informative Vibrotactile Displays to Support Attention and Task Management in Anesthesiology, Doctor of Philosophy Dissertation, University of Michigan (2010)

Using Sound Patterns to Enhance Directional Sound for Emergency Route Guidance

Tom Plocher[1], Zhaoxia Janet Jin[2], and Foong-Yeen Donny Chan[2]

[1] Honeywell International, Automation and Control Solutions, Advanced Technology Laboratory, 1985 Douglas Drive, Golden Valley, MN 55422 USA
[2] Honeywell Technology Solutions-China, 430 Li Bing Rd, Zhangjiang Hi-Tech Park, Shanghai, 201203 China

Abstract. Broadband sound emitted by digital sounder devices has been shown be a useful means for guiding building occupants to the correct emergency, particularly under conditions of darkness and obscuration from smoke. Standard practice has been to place a single sounder by the emergency exit door to function as an "auditory exit sign." The current paper reports on studies conducted in full scale building environments that investigated ways in which multiple sounders could be used along a route in a building to provide emergency route guidance to building occupants. The relative effectiveness of various patterns of sounder activation patterns is described. One such pattern that uses sequential activation of sounders together with number of sound pulses to encode sounder position along the evacuation route is particularly promising.

Keywords: directional sound, guidance, audio, evacuation, buildings.

1 Introduction

Withington [1,2,3] was the first to demonstrate the practical use of directional broadband sound as a means to guide building occupants to the correct emergency exit. Withington used a device that generated a locating sound comprised of a majority of the frequencies in the human hearing range emitted simultaneously. The result is a "shhhhhh" sound that is readily localized by humans. Withington's concept placed a single directional sounder device at each exit. Sounding the device attracted people in the vicinity to the nearest exit. Thus, essentially it was an audio version of the visual Exit signs commonly seen above exit doors in North American buildings. However, directional sound has been shown to be effective even in conditions of darkness or heavy smoke obscuration, conditions in which Exit signs are not visible or useful.

Withington's initial concept used single directional sounders placed at exit doors and were aimed at assisting evacuations in relatively small building spaces. However, more advanced concepts, involving many directional sounders arrayed at 10-15 meter intervals and emitting sound in a coordinated manner have been proposed as a means for actually guiding building occupants down a safe route of evacuation. These advanced route guidance concepts are promising, particularly when combined with new models for predicting the spread of smoke associated with a fire and hence, identifying safe evacuation routes.

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 329–337, 2011.

The research reported here was conducted to understand how to enhance current practice in coordinating multiple directional sounders to create sound patterns that are effective for route guidance in both small and large building spaces.

2 Pilot Tests

Initially, pilot tests were conducted in order to better understand how humans recognize and respond to directional sound patterns. The issues identified by the pilot tests then became the focus of a more formal and larger scale experiment.

2.1 Method

Directional Sound Device. All of the pilot tests and the experiment reported here used a directional sounder device called *ExitPoint*™, marketed by Honeywell International, Inc. This device has an integral audio amplifier that produces a pulsating sound consisting of broadband low, mid, and high range sounds. In addition to the broadband noise, the sounder is capable of playing an alert message in the form of a recorded voice message or other audible signals. The device, shown in Figure 1, comes from the factory pre-programmed with the four alternative pulse patterns shown in Table 1. The concept of use is to array the sounders along an evacuation route, setting each sounder to a different pulse rate, from Slow for the first sounder along the route, to Mid2 to Mid1, and finally to a Fast rate at the end of the route, e.g. the emergency exit door. Examples of simple sounder arrays are shown in Figures 2 and 3.

Table 1. Original pulse patterns

	Duration ON (msec)	**Duration OFF (msec)**
Slow	300	400
Mid2	225	200
Mid1	225	100
Fast	150	50

Fig. 1. *ExitPoint*™ sounder device

Fig. 2. Straight line sounder arrangement

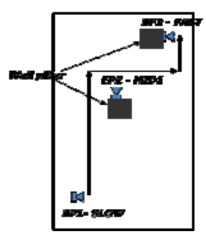


Fig. 3. "S" Shape Arrangement with sounders placed at turns in corridor

Our first pilot test observations were conducted in a public cafeteria of dimensions 50m x 15m. It was a highly reverberant environment, with tiled floor and a combination of glass brick walls. The ambient environmental noise measured at 67dB owing to people at work and air-conditioning noise. The sounders were set to emit sound at ¼ watt power and were placed at regular 10m intervals along the routes. Each sounder was set at a different pulse rate and, in terms of pulse rate, were arrayed from Slow to Medium 2 to Medium 1 to Fast, along the route. In some trials, all sounders were activated simultaneously throughout the trial. In other trials, only two adjacent sounders were activated simultaneously at any given time. On some trials, subjects were instructed to walk in a straight line from Slow to Fast as in Figure 2. In other trials, they were instructed to walk from Slow to Fast along a route laid out in an "S" shape as in Figure 3. Subjects' behavior was recorded on video and they were debriefed about their experience after the trials. Six subjects participated in these initial pilot trials.

These initial limited pilot trials suggested that mutual interference between adjacent sounders potentially was a problem and deserved further investigation. The initial trials also suggested that we needed a larger and more complicated test environment to represent a more typical environment in which sounders would be used. Thus, a second series of pilot test trials were conducted to discover how to make each individual sounder easier to discriminate from the sounders adjacent to it. Two deployment concepts were set up and compared:

- **Single Exit Sounder-** A single sounder set to Fast and also emitting an audible message, "Exit Here", was placed at the exit door at the end of the evacuation route. No other sounders were used.
- **"Sounder All @ Once"**- In this concept, four sounders were placed along the route from Slow to Mid2 to Mid1 to Fast at the exit door, activated simultaneously. In some trials, the Off component of the pulse rate for the Mid2 sounder was changed from its original factory setting of 200msec to a new value of 300msec in an attempt to make it more distinct from the Mid1 sounder and better guide the subjects through the middle portion of the route. As in the first concept, the Fast sounder at the exit door also emitted an audible "Exit Here" message.

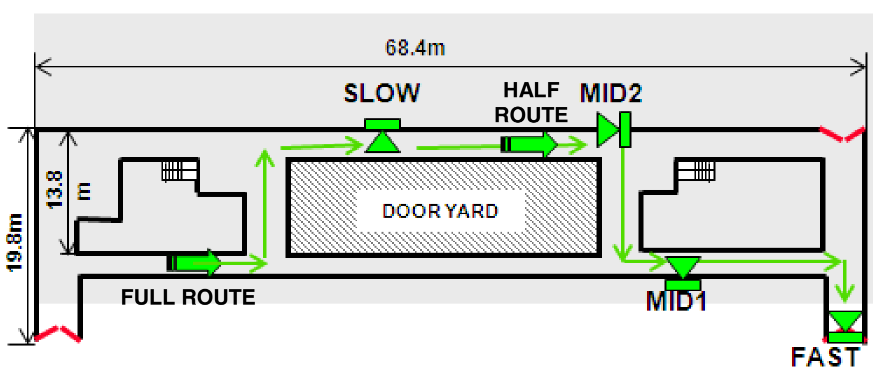


Fig. 4. Test environment for second pilot test

Pilot Test Protocol. Twenty-four college students with normal hearing participated. The larger and more complicated test environment shown in Figure 4 was used, a single floor of a university class building with typical concrete and glass construction. Subjects were instructed that their task was to move along an evacuation route to an exit door by following the sounders. On each trial, the subject started from either of the two starting points shown in Figure 4 by the bold green arrows. One starting point, noted as “Full Route” in the figure, was at the very beginning of the route and well ahead of the first sounder in the sequence. The subject’s task was to identify and move to the first sounder (the Slow one) and then on along the route. The distance of travel along the full route was 83.5 meters. The other or “Half Route” starting point shown in the figure was in between the first and second sounders in the sequence. In these half route trials, the subject’s initial task was slightly different. He or she first had to determine which direction to go, move toward the correct sounder (the Mid2 in this case) and then on along the route. This simulated the very real case that an evacuee might not always evacuate from the beginning of the route, but might just as readily find himself or herself starting somewhere in the middle and immediately facing a choice of which direction to head. The distance of travel along the half route was 46.8 meters.

Subjects wore half-blindfolds to simulate an environment filled with smoke, limiting visibility to less than 1 meter. After donning the blindfold before each trial started, the subject was lead around the corridors of the space in order to reduce any spatial memory of building layout. Barriers with alarm flags were used to show "danger here!" and prevent participants from wandering outside the planned test route.

2.2 Pilot Test Findings

Our pilot test observations highlighted a number of issues and potential deployment concepts for enhancing the effectiveness of the sounders for route guidance:

- **Single Sounders**- Single sounders were easy to hear and locate in the cafeteria setting of the first pilot test in which the user was within audible range of the single sounder all the time. In this situation the single sounder provided effective guidance. Single sounders were less useful in the larger university building setting of the second pilot test. In this larger setting, of course, a single exit sounder cannot provide selective guidance when alternate routes to the exit are possible. Also, users reported that they felt confused and scared when they were outside of the single sounder’s audible range and couldn't hear the voice message "Exit Here".
- **Sounder All At Once Concept-** This concept was able to provide route guidance in the large classroom/office building environment of the second pilot test. Subjects could hear guidance cues wherever they were along the evacuation route. Subjects reported that it was especially useful in helping them make the correct turn at choice points, where paths crossed each other. However, the pilot tests also revealed some issues with this concept:
  - **"Single vicinity zone effect"**- When subjects were very close to a particular sounder, they could hear only that sounder and not the adjacent ones, causing hesitation about which direction to go next.
  - **Mutual interference between sounders**- Subjects had difficulty discriminating between pulse rates when all four sounders were turned on

simultaneously and they were in between sounders along the route. They often paused to listen and determine which sounder was dominant and the appropriate one to follow. This hesitation was reflected in their times to traverse the evacuation route, which tended to be slightly longer than when guided by a single sounder. The tradeoff of their hesitation was more correct decisions about the path compared to the single exit sounder trials. Activating only two sounders at a time made each more distinct and appeared to be useful for route guidance.

- **Meaning of Fast versus Slow**- Some participants confused the meaning of Slow and Fast pulse rates and mistakenly thought that Slow meant safe. Training eliminated this problem.
- **Recognizing certain pulse rates-** Subjects easily recognized the Slow and Fast pulse rates, but reported difficulty recognizing the Mid2 and Mid1 rates.
- **Discriminating between adjacent pulse rates**- Subjects readily discriminated between Fast and Slow pulse rates and between Slow and Mid1. But they had difficulty discriminating between sounders that were closely related in pulse rate and adjacent along the route, e.g. between Slow and Mid2 rates, between Mid1 and Fast rates, and between Mid2 and Mid1. Changing the Off component of the pulse rate for the Mid2 sounder from its factory setting of 200msec to a new value of 300msec made it more distinct from the Mid1 sounder and provided better guidance to the subjects through the middle portion of the route.
- **"Fast" sounder at exit door-** A sounder at the exit door, set to Fast and sounding the audible message, "Exit Here" reduced the subject's confusion as long as it was always within clear audible range.

4 Experiment

A more formal experiment was conducted to compare the effectiveness of three sounder arrangements: 1) the Single Sounder concept, 2) the Sounder All @ Once concept using the modified wider pulse rate, and 3) a new concept called, Sequential Loop with Redundant Cues. In this new concept, as in the "Sounder All @ Once" concept, the four sounders were placed along the route from Slow to Mid2 to Mid1 to Fast at the exit door. However, this concept differed in three ways. First, in order to reduce mutual interference, the sounders were activated one at a time, in sequence, not simultaneously. When the sequence completed, it returned to the Slow sounder at the beginning and repeated the sequence, as in a loop. Second, the Slow, Mid2 and Mid 1 sounders were set to pulse or "sh" for 250 msec followed by a pause of a minimum of 200 msec. The Fast sounder was also set to pulse or "Sh" for 250 msec but each pulse was followed by a pause of only 100 msec. It was perceived as pulsing faster. Third, the pulse rates were manipulated so as to provide a redundant audio cue indicating position. Figure 5 illustrates this. Each sounder emitted a unique number of "sh" pulses to indicate its position in the sequence. Thus, the Slow sounder always emitted only one "sh", the Mid2 emitted two "sh" sounds, Mid1 emitted three "sh" sounds, and Fast emitted four "sh" sounds followed by the "Exit Here" message. After each sounder emitted its unique number of pulses, it remained silent until the sequence was complete and looped back to it for another iteration.

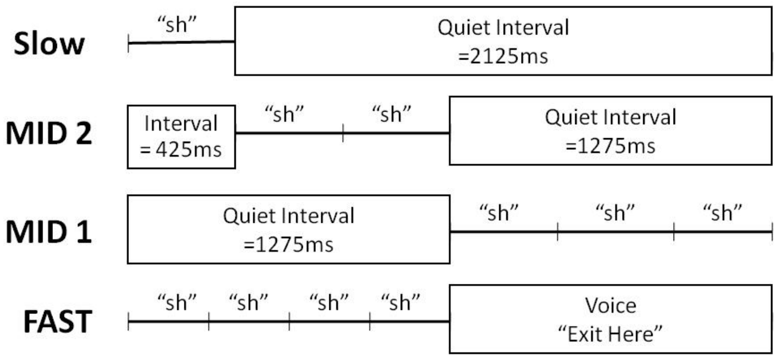


Fig. 5. Sequential loop pattern with redundant cues

4.1 Method

Design. A Sound Pattern X Route Within Subjects design was used. There were three sound patterns (Single Sounder, "Sounder All @ Once", Sequential Loop with Redundant Cues) and two routes (Full and Half). The order of treatments for each subject was counterbalanced. Dependent measures were completion (evacuation) time and error rate. Time of Completion was defined as time consumed by moving from the starting location to the sounder at the exit door. Error rate was defined in terms of the number of decision points where the subject had to go in one direction or another along the hallway or turn a corner. The "Full Route" had six such decision points. The "Half Route" had four. An error was recorded whenever the subject was observed to have walked in the wrong (unintended) direction, turned the wrong way or hit a barrier. Post-trial debriefings were done to explore with subjects the strategies they had used to find the exit and also their preferences for and acceptance of the three different sounder guidance concepts.

Test Protocol. Twenty college students with normal hearing participated in the experiment. The test environment was the same as that used in the second pilot test and shown in Figure 4. The test protocol was also the same with the exception that subjects participated in trials using three sounder concepts rather than two.

Table 2. Total number of decision point errors in route following with various sounder patterns.

Sounder Patterns	Route	
	Half Route	Full Route
Sounders All @ Once	5	6
Sequential Loop	2	3
Single Sounder	12	13

*24 trials for each cell.

4.2 Results

Errors. Table 2 shows the frequency of decision point errors across 24 trials in each condition. The frequency of errors was virtually the same in the Half Route and Full Route trials, so Chi Square analysis was performed on data collapsed over Route. A Chi Square analysis showed a significant effect of Sound Pattern on decision point errors ($x^2 = 15.4$, df = 2, $p < 0.00045$).

Time of Completion. Table 3 shows the results for time to complete each egress trial with the various sounder patterns and with the Full Route and Half Route starting points. Analysis of Variance found a significant effect of Sounder Pattern ($F = 4.764$, df=46, $p < 0.014$) and of Route ($F = 45.544$, df = 23, $p < 0.0001$). The interaction between Sounder Pattern and Route was not significant. Multiple comparisons between Sounder Patterns collapsed across route showed that time to complete egress was faster with both Single Sounder and Sequential Loop patterns than with the Sounder All @ Once pattern. Time of egress did not differ significantly between Single Sounder and Sequential Loop.

Table 3. Mean time (sec.) to complete a single egress trial with various sounder patterns and two different starting points.

Sounder Patterns	**Route**	
	Half Route	**Full Route**
Sounders All @ Once	69.34 (28.80)	95.34 (29.62)
Sequential Loop	53.81 (13.65)	88.88 (17.25)
Single Sounder	53.00 (8.83)	94.26 (26.59)

*24 trials for each cell.

Debriefing Results. At the conclusion of their test trials, subjects were debriefed and asked to provide ratings in response to several questions related to their experiences using the sounders for guidance and their preferences.

- **Sounder pattern preference-** Subjects' ranking of the three sounder patterns in order of preference revealed a significant preference for Sequential Loop over both the Sounders All @ Once and Single Sounder patterns ($p < 0.01$ for both comparisons by Wilcoxon Test) and for Sounder All @ Once over the Single Sounder pattern ($p < 0.05$ by Wilcoxon Test).
- **Egress strategy-** Subjects were asked if they completely followed the sounder guidance or only partially followed the sounds, relying on other cues such as feeling the walls, barriers, etc. 83% indicated that they completely followed the sounder guidance in their Sequential Loop trials, compared to 79% for the Sounder All @ Once trials and 33% for the Single Sounder trials.

5 Discussion

The results of the pilot studies and formal experiment suggest that using broadband sound emitted in carefully designed patterns can indeed be an effective medium for emergency route guidance in low visibility conditions. All three patterns investigated here, Single Sounder, Sounders All@ Once, and Sequential Loop with Redundant Cues, were able to guide subjects through evacuation routes laid out in various ways. Subjects committed significantly more errors in trials with the Single Sounder and also expressed their lowest preference for the Single Sounder. Subjects had to be within audible range of the single sounder for it to be useful, and even then, sometimes had difficulty localizing it and following it. So, while it can be effective under some conditions, the Single Sounder may be best suited as a redundant cue emitted in combination with one of the multiple sounder patterns.

The results also showed that, in the multi-sounder concepts, discriminating between adjacent pulse rates is key to providing clarity of guidance. Changing the Off component of the pulse rate for the Mid2 sounder from its factory setting of 200msec to a new value of 300msec made it more distinct from the Mid1 sounder and provided better guidance to the subjects through the middle portion of the route. However, even with this change, egress times in the half route trials for the Sounder All @ Once pattern were longer than one would expect if one were simply walking twice the distance. Subjects reported that this was due to some initial hesitation about which of the two adjacent sounders to follow. This was less an issue when a clear and non-overlapping sequence of sounds was provided. Even greater pulse rate separation than was used in our experiment would likely mitigate this problem and make the Sounder All@ Once concept even more effective.

The Sequential Loop With Redundant Cues pattern was the most preferred method for guidance and resulted in the fewest errors at decision points. It's sequential pattern eliminated interference from adjacent sounders and effectively "pointed the way." Also, it highlighted the potential for improving route guidance by adding redundant cues, in this case, using the number of pulses at each sounder to encode its position in the sequence. Future research should more fully explore such concepts.

This was an initial investigation of using sounders in combinations to create patterns for emergency route guidance. Many questions remain. For example, using verbal cues or messages along the route in combination with the broadband sound might reduce some of the ambiguity of sound alone. Verbal cues will have to be carefully engineered to minimize interference with the broadband sound being emitted. Another question that must eventually be addressed is that of route adaptation. In the future, using computer models, we will have the ability to predict smoke and heat propagation from a fire, and classify the safety of evacuation routes according to those predictions. That creates the possibility of dynamically adapting routes to the conditions of smoke and heat and ensuring that evacuees are always guided along safe routes. But how is a change in route communicated to them? Future research will have to address that.

References

1. Withington, D.J.: The use of directional sound to improve the safety of auditory warnings. In: Proceedings of the XIVth Triennial Congress of the International Ergonomics Association & 44th Annual Meeting of the Human Factors and Ergonomics Society, San Diego, pp. 726–729 (2000)
2. Withington, D.J.: Faster evacuation from ferries with sound beacons. Fire 30 (2000)
3. Withington, D.J., Smith, S.: "White Noise" can help save stricken firefighters. Fire 27 (2000)

A Knowledge Elicitation Study for a Speech Enabled GIS to Handle Vagueness in Communication

Hongmei Wang

Northern Kentucky University, Department of Computer Science, Nunn Drive,
Highland Heights, KY 41099, USA
wangh1@nku.edu

Abstract. One of challenges toward development of usable speech enabled Geographical Information Systems (GIS) is how to handle vagueness that naturally exist in human-GIS communication. The meaning of some spatial concepts are not only fuzzy, but also context dependent. To enable the GIS to handle the vagueness problem, in particular, the context-dependency sub-problem, we propose to design a collaborative speech enabled GIS, which can emulate a human GIS operator's role and handle the vagueness problem in communication through collaborative dialogues. To emulate a human GIS operator's role, the GIS must have knowledge corresponding to a human GIS operator's knowledge involved in handling the vagueness problem. This paper describes a knowledge elicitation study that we conducted to elicit human GIS operators' knowledge about how to handle the vagueness problem through collaborative dialogues. A speech enabled GIS, Dave_G, incorporates part of the study results. This system is able to handle the vagueness problem through various collaborative dialogues.

Keywords: GIS, Knowledge elicitation study, Vagueness problems, Human-GIS Communication, Collaborative dialogue strategies.

1 Introduction

Geographic Information Systems (GIS) are computer systems for capturing, storing, querying, analyzing, and displaying geographically referenced data [1], such as map layers of various spatial features (e. g. land parcels, roads, lakes, cities etc.). They have been broadly applied in various fields that use spatial data since the 1960s, for instance, natural resource management and assessment, criminal incident analysis, crisis management, automobile GPS, web-based map systems, etc.

To improve the system usability and support broader use of the system, the GIS community has spent much effort on designing transparent interfaces. One line of the efforts is to develop natural interfaces, e. g. speech and gesture enabled interfaces, to GIS [2-6]. There have been some experimental natural interfaces for geographic information use, such as "*Put-that-There*" [7], CUBISON [8], *QuickSet* [9-11], *Sketch and Talk* [12], *iMap* [13], and *Dave_G* [14].

Most of the existing experimental systems use speech as one of interaction modalities in communication with the user. To develop a usable sppech enabled GIS, one of

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 338–345, 2011.

the challenges that must be addressed is how to semantically interpret a speech request involving spatial concepts. The user's natural language requests can be incomplete, ambiguous, vague, and inconsistent [15-18]. The meaning of a vague linguist term representing a spatial concept is fuzzy, that is, its meaning does not have a clear boundary. Its meaning is also context dependent, that is, its meaning is different in different contexts. This study focuses on handling the vagueness problem in human-GIS natural language communication, in particular, the context-dependency sub-problem.

Human communicators can usually successfully communicate vague spatial concepts with each other through collaborative dialogues. For example, in an Emergency Operation Center for the hurricane crisis management in Florida, a crisis manager who does not know to operate GIS can collaborate with an expert GIS operator and can still get desired spatial information from the GIS although their natural language communication may involve vagueness. Success in collaborative human-human communication drives us to design a speech enabled GIS which can handle the vagueness problem by emulating a human GIS operator's role. The goal of this study is to collect human knowledge about how human communicators, in particular, human expert GIS operators, understand the meaning of a vague spatial concept in different contexts and reduce vagueness in communication through collaborative dialogues.

Knowledge elicitation is the process of collecting information that is thought to be relevant to desired knowledge from a human resource of knowledge [19]. There are various types of knowledge elicitation techniques [20-23]. The observations and interviews are often used for early stage of system design, such as, conceptualizing the problem domain. They are direct methods of watching experts and talking to them. In this study, we took these two techniques to collect human communicators' knowledge involved in handling vagueness in communication. This paper describes design, the data collection process, and preliminary findings of the knowledge elicitation study.

2 Research Design

Our research questions in the knowledge elicitation study are:

- 1) How do human communicators understand the meaning of a vague spatial concept in communication?
- 2) What kinds of collaborative dialogue strategies can be used by human communicators to reduce vagueness in communication of spatial concepts?
- 3) How does the human GIS operator reason when the user speaker's request involves a vague spatial concept?

The first knowledge elicitation technique that we planned to use is participant observation. We planned to invite pairs of GIS experts and non-expert GIS users to work together on a set of tasks (see Table 1) that involve communication of a vague spatial concept, *near*, and use of GIS commands. Through observing their collaboration process, we expected to answer the research question 2 and observe the communication between pairs of human communicators, in particular, collaborative dialogue strategies used to reduce vagueness in communication.

The eight tasks are about getting a map near a reference location. The maps are required to fit for different contexts (Table 1). They have some contextual factors in common, such as the goal, time frequency/period and bigger spatial contexts, also have different contextual factors, such as transportation mode and reference locations.

Table 1. Eight Tasks in Participant Observation

Task No.	Task Content	Common Contextual Information	Different Contextual Information	
			Transportation mode	Reference location
1	Show a map showing grocery stores *near* a home location.	Goal: food shopping; Time: Weekly; Space: In a big city like New York	By driving a car	Apartment A
2			by walk	Apartment A
3			By driving a car	Apartment B
4			by walk	Apartment B
5	Show a map showing cities *near* a home city.	Goal: Vacation; Time: two day weekend	By driving a car	City A
6			By air	City A
7			By driving a car	City B
8			By air	City B

Near is a typical example used to illustrate the vagueness problem in communication of spatial concepts [15, 24-27]. We expected that using this concept in the tasks would lead to vagueness in communications between each pair of participants. We hoped that the vagueness problem would further stimulate the participants to use collaborative dialogues.

The second knowledge elicitation technique is interview. We also planned to interview the participants after the observation and expected to answer the research question 3 by collecting data about the human GIS operators' reasoning process underlying their collaborative communication with users during the observation process. We also expected to answer the question 2 about their different understandings of the same vague spatial concept in different communication contexts.

3 Data Collection

Four pairs of GIS expert operator participants and non-expert user participants participated in our study. They were all university staff or graduate students. All of them spoke in English fluently. This section details the entire knowledge elicitation data collection process.

3.1 Introduction Session

After a pair of GIS expert operator participant and non-expert user participant came to our experiment location (a computer lab), we first distributed consent forms to them and collected the forms back when they were done with reading and signing the forms.

Consequently, to the user participant, we introduced basic concepts of GIS, such as its definition, major functions and applications. We also explained the eight tasks (Table 1) to be completed later with the GIS expert participant together. We required the user participant to keep the context specified for each task in mind and express each request as naturally as possible to the operator participant.

Finally, to the operator participant, we demonstrated major GIS functions, in particular, those that would be used to complete the required tasks. The demonstration was on a laptop to be used by the operator later to help the user participant. It was installed with GIS software, ArcView 3.1. We also required the operator participant to practice using the laptop and the GIS software, in particular, those major GIS functions.

3.2 Observation Session

The pair of participants started to work on the eight tasks together after the introduction session. The user participant read requirements for each of the eight tasks at first, and then talked to the operator participant for each request. The operator participant helped the user participant to generate a map for each task. They often had several round of collaborative dialogues for most of the tasks.

When the pair of participants collaborated with each other on these tasks, we recorded the entire collaboration process via a video camcorder. The researcher focused on observing and taking written notes on their collaborative dialogue strategies and major their facial expressions, in particular, when they stopped talking or working and were hesitating for something. Those collaborative dialogue strategies used and stopping would lead to questions and discussion in the later interview.

3.3 Interview Session

We interviewed the participants right after the observation session so that they would still remember what happened in the observation session and what they had thought in order to handle the vagueness problem. The interview process was recorded as digital audio data and written notes.

At first, we gave a short interview to the user participant. The participant was asked with questions about how they understood the meaning of the vague spatial concept near in the tasks 1 to 4 and how they did in the group of tasks 5 -8. We also asked the participant to list all possible collaborative dialogue strategies that he/she thought could help to handle the vagueness problem and order them.

We interviewed the operator participant consequently. In addition to the questions asked for the user participant, the questions to the operator participant focused on his/her reasoning process involved in handling the vagueness problem. We asked the reason why some collaborative dialogue strategies was used in some of the tasks and the participant's mental belief on the parameter that was instantiated by the vague spatial concept *near* and the execution result of GIS commands that used the parameter value instantiated from *near*. We also asked why they stopped working or hesitated for a while during the observation session if they did so.

4 Findings

The data that we collected all are qualitative. So, we take an interpretative reading method to read and analyze the data collected. This section summarizes the findings from our data analysis.

4.1 Understanding of a Vague Spatial Concept

The meaning of a vague spatial concept is understood in a context where the map is used. For example, the GIS operator participants understood the meaning of the vague spatial concept *near* in communication by imaging the situation that the requested map was used for. One operator participant showed a map of grocery stores that he thought were near to the user after receiving the request from the user. He explained how he estimated near during the operation in the interview. He was thinking how he was going to shop, such as driving a car from the specified apartment to those grocery stores. Then, he assumed that this was the user's reason to ask for the map.

4.2 Collaborative Dialogue Strategies

We find five major collaborative dialog strategies that the GIS operator can take to handle the vagueness problem:

- Strategy 1: The GIS operator shows a map result with a parameter value estimated from a vague spatial concept, and then asks the user for comments on the map result;
- Strategy 2: The operator directly tells the user an estimated meaning of the vague spatial concept by the operator and asks the user for feedback on the estimated meaning directly;
- Strategy 3: The GIS operator asks the user for some major contextual factors that influence the meaning of the vague spatial concept to have a better understanding of the concept;
- Strategy 4: The GIS operator directly asks the user for the meaning of the vague spatial concept that the user intends and then uses it in the GIS operation;
- Strategy 5: The GIS operator asks the user for modifying comments on an existing estimated meaning of a vague spatial concept.

4.3 Reasoning Process

When the GIS operator receives the request involving a vague spatial concept, he/she needs to make a decision about what to do next, e. g. taking one of the five collaborative dialogue strategies described above. The recognition-primed decision (RPD) model [28, 29] explains human's decision making process. To explain the GIS operator's decision-making process involved in handling the vagueness problem, we analyze our interview data and summarize major features of a computational RPD model that describes the GIS operator's decision making process. The results are given in Table 2.

Table 2. Major Features of a Computational RPD Model

<table>
<tr><td colspan="2">Key Cues
(to identify the problem)</td><td rowspan="2">Goals</td><td rowspan="2">Expectancies</td><td rowspan="2">Course of Actions</td></tr>
<tr><td>Common</td><td>Different</td></tr>
<tr><td rowspan="5">The user utterance is clear;
There is a GIS command available for the request;
A spatial concept expressed in the user's utterance is vague;
The user has not provided an exact meaning of the vague concept.</td><td rowspan="2">Some contextual information is provided by the user or assumed by the operator. The operator has knowledge about how to estimate a meaning of the vague spatial concept based on context information.</td><td rowspan="5">Instantiate a para in the GIS command from the vague concept</td><td>The user is able to provide comments on the estimated meaning of the vague spatial concept shown on the map</td><td>Strategy 1</td></tr>
<tr><td>The user is able to provide comments on the explicitly expressed estimated meaning of the vague spatial concept.</td><td>Strategy 2</td></tr>
<tr><td>Part of major context information has not been shared by the user.</td><td>The user is able to provide such context information.</td><td>Strategy 3</td></tr>
<tr><td></td><td>The user is able to provide an exact meaning of the vague spatial concept.</td><td>Strategy 4</td></tr>
<tr><td>An existing estimated meaning of the vague spatial concept has been provided to the user.</td><td>The user is able to provide modifying comments on the existing estimated meaning of the vague spatial concept.</td><td>Strategy 5</td></tr>
</table>

5 Conclusion

This paper describes a knowledge elicitation study that we conducted to help a speech enabled GIS to handle the vagueness problem in human-GIS communication. The preliminary findings about how human communicators understand the meaning of a vague spatial concept in communication help us to better understand the vagueness problem in human-GIS communication and how the GIS can understand the meaning of a vague spatial concept in the communication context. The findings about various collaborative dialogue strategies and the human GIS operator's decision-making process help us to design intelligent speech enabled GIS which can handle the vagueness problem through collaborative dialogues. These findings are also useful for design of other speech enabled information systems to handle vagueness in human-computer communication.

Acknowledgments. This material is based upon work supported by the National Science Foundation under Grants No. BCS-0113030, and EIA-0306845. This work is also supported by Northern Kentucky University Faculty Summer Fellowship and Northern Kentucky University Faculty Project Grant. The author would like to thank Dr. Guoray Cai, Dr. Alan MacEachren and Dr. Michael McNeese for their advising to her work during her stay at Pennsylvania State University.

References

1. Chang, K.-t.: Introduction to Geographic Information Systems. McGraw-Hill, New York (2010)
2. Frank, A.U., Mark, D.M.: Language issues for GIS. In: Macguire, D., Goodchild, M.F., Rhind, D. (eds.) Geographical Information Systems: Principles and Applications, pp. 147–163. Wiley, New York (1991)
3. Mark, D.M., Svorou, S., Zubin, D.: Spatial terms and spatial concepts: Geographic, cognitive, and linguistic perspectives. In: Proceedings of International Geographic Information Systems (IGIS) Symposium: The Research Adgenda, vol. 2, pp. 101–112 (1987)
4. Florence, J., Hornsby, K., Egenhofer, M.J.: The GIS Wallboard: interactions with spatial information on large-scale display. In: Kraak, M.-J., Molenaar, M. (eds.) Seventh International Symposium on Spatial Data Handling (SDH 1996), pp. 8A. 1–15 (1996)
5. Mark, D.M., Frank, A.U.: NCGIA Initiative 2, "Languages of Spatial Relations". Closing Report.: National Center for Geographic Information and Analysis, Santa Barbara, CA (1992)
6. Mark, D.M., Frank, A.U.: User interfaces for Geographic Information Systems: report on the specialist meeting. National Center for Geographic Information and Analysis (1992)
7. Bolt, R.A.: Put-that-there: voice and gesture at the graphics interface. In: Proceedings of the 7th Annual Conference on Computer Graphics and Interactive Techniques, pp. 262–270 (1980)
8. Neal, J.G., Thielman, C.Y., Funke, D.J., Byoun, J.S.: Multi-modal output composition for human-computer dialogues. In: Antonisse, H.J., Benoit, J.W., Silverman, B.G. (eds.) Proceedings of the 1989 AI Systems in Government Conference, pp. 250–257 (1989)
9. Cohen, P., Dalrymple, M., Moran, D., Pereira, F.: Synergistic use of direct manipulation and natural language. In: CHI 1989, pp. 227–234. ACM/Addison Wesley (1989)
10. Oviatt, S.L.: Pen/voice: Complementary multimodal communication. In: Proceeding of Speech Technology 1992, pp. 238–241 (1992)
11. Oviatt, S.L.: Multimodal interfaces for dynamic interactive maps. In: Proceedings of the Conference on Human Factors in Computing Systems (CHI 1996), pp. 95–102 (1996)
12. Egenhofer, M.J.: Multi-modal spatial querying. In: Kraak, J.M., Molenaar, M. (eds.) Advances in GIS Research II (Proceedings of The Seventh International Symposium on Spatial Data Handling), pp. 785–799. Taylor & Francis, London (1996)

13. Sharma, R., Yeasin, M., Krahnstoever, N., Rauchert, I., Cai, G., Brewer, I., MacEachren, A.M., Sengupta, K.: Speech-gesture driven multimodal interfaces for crisis management. Proceedings of the IEEE 91, 1327–1354 (2003)
14. MacEachren, A.M., Cai, G., Sharma, R., Rauschert, I., Brewer, I., Bolelli, L., Shaparenko, B., Fuhrmann, S., Wang, H.: Enabling Collaborative GeoInformation Access and Decision-Making Through a Natural, Multimodal Interface. International Journal of Geographical Information Science 19, 293–317 (2005)
15. Wang, F.: Handling Grammatical Errors, Ambiguity and Impreciseness in GIS Natural Language Queries. Transactions in GIS 7, 103–121 (2003)
16. Bosc, P., Prade, H.: An introduction to fuzzy set and possibility theory based approaches to the treatment of uncertainty and imprecision in database management systems. In: Proceedings of the 2nd Workshop on Uncertainty Management in Information Systems: From Needs to Solutions (1993)
17. Owei, V.: An intelligent approach to handling imperfect information in concept-based natural language queries. ACM Transactions on Information Systems 20, 291–328 (2002)
18. Robinson, V.B., Thongs, D., Blaze, M.: Machine acquisition and representation of natural langauge concepts for geographic information retrieval. In: Proceedings of 16th Annual Pittsburgh Conference, vol. 1, pp. 161–166 (1985)
19. Cooke, N.J.: Varieties of knowledge elicitation techniques. International Journal of Human-Computer Studies 41, 801–849 (1994)
20. Boose, J.H., Bradshaw, J.M.: Expertise transfer and complex problems: using Aquinas as a knowledge-acquisition workbench for knolwedge-based systems. International Journal of Man-Machine Studies 26, 3–28 (1987)
21. Cordingley, E.S.: Knowledge elicitation techniques for knowledge-based systems. In: Diaper, D. (ed.) Knowledge Elicitation: Principles, Techniques and Applications. Ellis Horwood Ltd., Chichester (1989)
22. Meyer, M.A., Paton, R.C.: Towards an analysis and classification of approaches to knowledge acquisition from examination of textual metaphor. Knowledge Acquisition 4, 347–369 (1992)
23. Olson, J.R., Biolsi, K.J.: Techniques for representing expert knowledge. In: Ericsson, K.A., Smith, J. (eds.) Toward a General Theory of expertise, pp. 240–285. Cambridge University Press, Cambirdge (1991)
24. Wang, F.: Towards a natural language user interface: An approach of fuzzy query. International Journal of Geographical Information Systems 8, 143–162 (1994)
25. Wang, F.: A fuzzy grammar and possibility theory-based natural language user interface for spatial queries. Fuzzy Sets and Systems 113, 147–159 (2000)
26. Yao, X., Thill, J.-C.: How Far Is Too Far? A Statistical Approach to Context-contingent Proximity Modeling. Transactions in GIS 9, 157–178 (2005)
27. Yao, X., Thill, J.-C.: Spatial queries with qualitative locations in spatial information systems. Computers, Environment and Urban Systems 30, 485–502 (2006)
28. Klein, G.A.: Recognition-primed decisions. In: Rouse, W.B. (ed.) Advances in Man-Machine Systems Research, pp. 47–92. JAI Press, Greenwich (1989)
29. Klein, G.A.: Recognition-Primed Decision Making. In: Klein, G.A. (ed.) How People Make Decisions, pp. 15–30. MIT Press, Cambridge (1998)

Believe What You Hear, Not What You See – Vision Interferes with Auditory Route Guidance in Complex Environment

Ying Wang[1], Huiting Zhang[2], Lu Yu[2], Kan Zhang[2], Xianghong Sun[2], and Thomas Plocher[3]

[1] Key Laboratory of Mental Health, Institute of Psychology, Chinese Academy of Sciences, China
[2] Institute of Psychology, Chinese Academy of Sciences, China
[3] Honeywell Labs, Minneapolis, Minnesota, USA
{zhanghuiting,yulu,zhangk,sunxh}@psych.ac.cn, wangying@psych.ac.cn, tom.plocher@honeywell.com

Abstract. Auditory route guidance has a potential use for sighted people who have to conduct emergent real-world task during navigation. Despite its affordance in assisting people in the absence of vision, it may receive interference from vision under normal visual condition. The present study tested the effect of vision on auditory route guidance using different display modes. Normal-sighted firefighters were instructed to navigate within a virtual building following auditory commands from a navigation aid, either under normal (high-visibility) or smoked (low-visibility) visual condition. Navigation in normal visual condition was faster but less accurate than that under low-visibility, and was characterized by unique walking patterns. Moreover, it resulted in worse spatial memory and less positive experience toward the system. These results suggest that the interaction mode of human and auditory route guidance system could be modified by vision. Clear visual inputs boost risk-taking behaviors in route following, which might lead to dangerous consequence in specific navigation tasks. Furthermore, the interference from vision was not restricted to specific display mode, indicating that it might be a general problem for auditory route guidance. As a challenging and primary human factor issue, it should attract more attention and caution in future research and design work.

Keywords: auditory route guidance, vision, firefighter, human-computer interaction, visual auditory interaction.

1 Introduction

As a species on the move, humans are adept in planning routes based on their spatial knowledge and the visual surroundings. However, this ability is often challenged by unfavorable conditions like a complex and unfamiliar environment, lack of valid visual cues, or engaging in a high demanding task. Personnel of specific occupations, for example the firefighters, may encounter all the mentioned problems in their work. Navigation aids with route guidance functions are urged for these people.

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 346–354, 2011.

Route guidance should have minimal interference with the interaction between the user and the real-world task, suggested by the previous work [1]. According to this principle, an auditory interface is a good candidate for navigation aids which are designed for sighted people with particular tasks. It does not require the users to take their eyes off the visual surroundings. Neither does it conflict with hand-operations. The affordance of auditory route guidance has been established through the past decades, mostly in aiding the visually impaired [2]. People with permanent or temporary visual impairment were used to evaluate the effects of different auditory display modes in terms of spatial perception [3] [4], route guidance [5] [6] [7] [8] [9], spatial learning [10] [11], and cognitive loads [12]. Nevertheless, from the other hand, navigation is essentially active and vision dominated for sighted people, yet the role of vision in auditory route guidance remains largely unknown.

Here we assessed the performance of an auditory route guidance system with a simulated firefighting task. Normal-sighted firefighters were recruited to accomplish the task with four auditory display modes either under high-visibility (normal) or low-visibility (smoked) condition. We aimed to investigate whether the effect of auditory route guidance would be affected by the presence of clear visual surroundings and whether the role of vision was dependent on display mode. We hypothesized that full visual inputs would enhance navigation performance to some extent compared with the low-visibility condition, but might also cause problems for the interaction between human and route guidance system. For instance, a conflict might arise when visual cues suggested a plausible alternative route different from the auditory commands.

2 Method

2.1 Participants

Twenty-four male firefighters with normal vision and hearing took part in the experiment. Ages ranged from 20 to 25. All had no experience with navigation aid in firefighting work. They were randomly assigned to high-visibility and low-visibility conditions in equal proportion.

2.2 Virtual Environment

The test environment was constructed based on the layout of a corporate office and laboratory building with a desktop virtual reality technique. The virtual building has irregular and complex internal structures (room, stairway etc.), monochrome walls and floors, and no other landmarks. In the high-visibility condition, which simulated the normal visual environment, the end of the route on which one was standing and all the non-obstructed structures in open areas could be clearly seen. In the low-visibility condition the environment was filled with grey smoke of uniform density, allowing participants to see only within several feet around them. We planned four routes through different parts of the building, each composed of nine segments joined by 8 waypoints. The lengths were 429.6ft, 483.9ft, 501.0ft, and 390.6ft. There were six fires randomly located along or several feet away from each planned route. See example of a floor plan with a predefined route and fires in Figure 1.

2.3 Auditory Guidance

A prototype of an indoor route guidance system was created. During the test, the guidance system guided the participants with auditory commands as they walked from one waypoint to another along a planned route from one end to the other. The participant's position within the virtual environment was recorded every 50ms. The course from the current position to the next waypoint was a derived straight line. The next waypoint was determined by two simultaneously satisfied rules: among the two that were closest to the participant's current position and nearer to the destination. Two types of commands were given respectively. Once the angle between the derived path and the participant's heading direction was larger than four degrees, turning commands were triggered. Otherwise (deviation less than 4 degrees), forward commands were given. A musical tone announced the arrival at each waypoint at the capture radius of 3.281 feet (1m).

For each visibility condition, there were four auditory display modes, with different combinations of turning and going-forward commands. The pure *speech* mode guided the user to turn (e.g., "turn right, 60 degrees") and go forward (e.g., "go forward, 10 meters") with non-spatialized verbal commands. The pure *3D-audio* mode used a "whoop" sound coming from the direction of the next waypoint to indicate turning, and a "ding" sound from ahead of the user for going forward. The *CombinedA* mode instructed turning with the spatialized `whoop' sound as in the pure 3D mode, while the subject was led going forward with 3D-speech (e.g., "go forward, 10 meters" as coming from ahead). The *CombinedB* mode, opposite to *CombinedA*, guided turning with 3D-speech (e.g., "turn right, 60 degrees" as coming from 60 degrees right) while directed going forward with the `ding' sound coming from ahead.

2.4 Apparatus and Parameters

The experiment was run on a DELL OptiPlex 745 Desktop with a 1.86GHZ Intel Core2 Duo Processor. The program was developed in C with Visual Studio 2005 and rendered in MOGRE (Managed OGRE). The 3D-audio effect was implemented by OpenAL and was created using Creative Sound Blaster X-Fi Surround 5.1 soundcard and LTB Magnum 5.1 AC97 Headphones. The speech commands were in clear female voices and synthesized by the "Fang Zheng Chang Ting" software (Founder Tech). All the 3D sounds were virtually 1 meter away from the user. The "whoop" sound lasted for 100ms with emission interval of 1 second. The "ding" sound was 900ms long. Its emission interval changed from 3 seconds to 0 second gradually from the position where the going forward commands were triggered to the next waypoint. All the sounds were displayed with a background white noise at its 1/20 intensity.

2.5 Procedure

The participants were instructed to rescue a person trapped by fire in a particular room inside the test building with the instructions of auditory guidance. They also needed to extinguish fires encountered on the way of rescue without running into them. They were informed that the navigation aid would generate a relatively short and safe route based on stored map but could not predict the number and locations of fires. Each participant finished four trials on four display modes under the same visibility condition, either with smoke or not. The two pure modes came first, followed by the two combined modes. The orders within each pair were counterbalanced. The combinations of routes and display modes were pseudo-randomized.

Before the first two formal trials, participants practiced the task within another virtual environment until they got familiar with the task. During the task, they continuously received auditory commands from the headsets and used “up", “down", “left", “right" keys to move forward, backward, turn left, and turn right. They pressed a “space" key to extinguish the fires detected on their way. Immediately after they reached the trapped person, a feedback sound announced success and a window popped up presenting the post-task questionnaire. In the questionnaire, they were asked to choose a route, from four standardized bird-view routes on a blank background, that represented the path that they had followed on that trial. Then they did five-point ratings on A) the difficulty of accomplishing the task under guidance, and B) their preferences for guidance mode.

3 Results

3.1 Task Completion

With the help of our navigation aid, all the participants were able to reach the destination efficiently in most of the trials, despite the fact that it was a very difficult task to find a designated location within an unfamiliar complex environment, particularly with limited visual cues. Two out of forty-eight trials (twelve participants by four modes) under the high-visibility condition and four under low visibility were failed. The failures were largely due to the computing limit of the navigation system. As a prototype, for algorithm economy, the system did not take account of the physical occlusions (walls) in route computing. While getting into a wrong enclosed region accidentally, the participant had to find his way back by bypassing the wall and ignoring the auditory commands temporarily. Failure to find the right exit back to the route potentially lead to a long detour and eventually quitting the trial. We excluded the failed trials from the following analysis and replaced each missing value by the mean of the other three trials from the same individual. Meanwhile, the mean numbers of extinguished fires were high in all the conditions (ranged from 4.3 to 5.8, with a total number of 6), suggesting that the participants had paid attention to the visual surroundings during route, following as required even when the visibility was low.

3.2 Navigation Time

We standardized the time of each trial by dividing the length of the planned route. As shown in Table 1, the participants spent less time to complete the task in all the four high-visibility conditions than in the low-visibility conditions. Mixed ANOVA showed significant main effects of both visibility: ($F_{(1,22)}$=26.47, p <.001) and display mode ($F_{(3, 66)}$=5.227, p=.003).

Table 1. Navigation performance measures of four display modes under high and low visibility conditions (SE in parenthesis)

	Standardized navigation time (sec/ft)		Number of passed waypoints		Absolute deviation from the planned route (ft)	
	High Vis.	High Vis.	Low Vis.	Low Vis.	High Vis.	Low Vis.
Speech	2.34 (.26)	3.95 (.26)	7.25 (.28)	8.00 (.00)	2.72 (.24)	2.26 (.24)
3D-audio	2.85 (.28)	4.62 (.28)	7.22 (.35)	8.00 (.00)	2.64 (.47)	2.54 (.47)
CombA	3.69 (.53)	5.42 (.53)	6.75 (.39)	7.83 (.11)	3.12 (.49)	2.04 (.49)
CombB	2.91 (.42)	4.03 (.42)	6.67 (.41)	7.89 (.09)	2.92 (.37)	1.65 (.37)

3.3 Accuracy of Navigation Path

Besides efficiency, assessing the degree of matching between the actually travelled paths and the planned routes also serves as an important index for navigation performance. It indicates to what extent the participants followed the short and safe route to the destination. Actually, in the firefighting scenario, priority should be given to safety over speed. Comparisons of route matching between the two visibility conditions yielded intriguing results. Generally, when visibility was low, the participants followed the commands along the planned route more accurately. However, in the normal visual environment (high-visibility), they did not follow the route exactly as were instructed. This resulted in some unique walking patterns.

Number of Passed Waypoints. The mean number of waypoints (maximum 8, excluding start and end) that the participants had arrived at in each condition is shown in Table 1. The participants reached almost all the waypoints under the low-visibility conditions. By contrast, at least one waypoint was missed on average in every high-visibility condition.

Absolute Deviation. We also calculated the mean absolute deviation of the recorded positions along the real walking path from the planned route for each trial. To exclude the large deviations which might be caused by active navigation, e.g., getting into a wrong compartment and trying to find the exit, deviation values more than 10 feet were set as 10. As shown in Table 1, route-following performance under the low-visibility condition was generally more accurate, or, deviated from the planned route less, compared with that under the high-visibility condition. Mixed ANOVA indicated a trend to significance for visibility ($F_{(1,22)} = 2.685$, $p = .112$). No other effect was found.

Walking Pattern. Figure 1 shows examples of recovered walking paths under the two visibility conditions. The travelled route under low-visibility was relatively smooth. However, when visual cues were easily accessible, there were some unique walking patterns. The participants sometimes ignored the commands and took a "shortcut" which was indicated by visual features (e.g., fire) far away from their current positions (see the courses from the 5^{th} to the 6^{th} waypoint counted from the right). Additionally, they occasionally moved ahead of the commands and made mistakes.

Fig. 1. Examples of recovered walking traces along one of the routes in high visibility (upper) and low-visibility (lower) conditions, both with the speech display mode. The thick solid line indicates the planned route. The thin dotted lines indicate walking traces of three participants. The participants started from the right end of the route to rescue the person that was at the left end of the route (dark icon). The icons along the route indicate fires. The circles at each waypoint cover the range of its capture radius. If the walking trace did not get through the circle around a waypoint, then this waypoint was regarded as a missing waypoint.

Two participants walked into dangerous open field and lost time on being guided back to the planned path (see the course from the waypoint on the lower-right corner).

3.4 Route Memory

In real work, the firefighters get out of a building along the way that they come in, for safety concerns. Gaining memory of the travelled route would facilitate the subsequent evacuation. Interestingly, although the participants in the high-visibility condition had more visual cues about the environment, their route memory was worse than that in the low-visibility condition ($F_{(1,22)}$=3.144, p=.09). The only two conditions that had average route recall accuracy above the chance level (0.25) were CombinedB (0.42) and speech (0.42) modes under low visibility. Recall accuracy of the other conditions range between 0.06 and 0.17.

3.5 Subjective Experience

The participants' overall evaluations on the navigation aid were quite positive, meanwhile distinguished between two visibility conditions (Fig. 2). The main effect of display mode was significant in the low-visibility condition on both task difficulty ($F_{(1,22)}$=3.484, p=.027) and preference ($F_{(1,22)}$=3.752, p=.02). The overall pattern was consistent with the behavioral observations. By contrast, display mode did not have significant impact on the subjective experience under high-visibility condition. We did not formally compare the absolute rating scores for the two visibility conditions since the two groups of participants rated based on unequal experience. However, it is interesting to note that the frequency of choosing "like" or "like very much" was generally high for the participants who only experienced the low-visibility environment, despite that they actually spent much more time on the task.

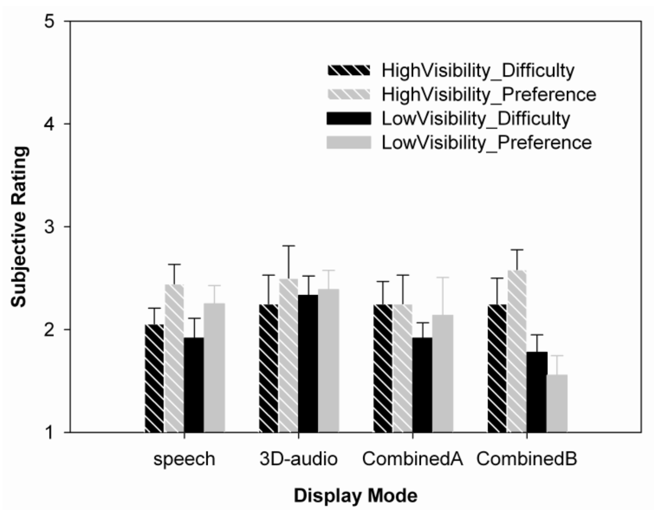


Fig. 2. Task Difficulty (1: very easy; 5: very hard) and Subjective Preference (1: like very much; 5: dislike very much) ratings on the four display modes under high- and low-visibility conditions.

4 Discussion

The major goal of the current study was to investigate how the availability of visual cues influences auditory route guidance. Intuitively, it would be easier to navigate within a normally visible environment than in the low-visibility condition. However, our results suggest otherwise when both visual cues and auditory guidance were provided. Within a clear visual environment (high visibility) the participants walked faster but tended to use auditory commands arbitrarily and sacrificed navigation accuracy. However, they were able to accomplish the same task more accurately within acceptable time when the visibility was low. The tradeoff suggests that availability of visual cues might modify the interaction mode between human and the route guidance system. In a normal visual condition, visual feedback confirms the progress of travelling instantaneously and may boost the participants' confidence on active navigation. As a result, they move faster and sometimes too fast, at the cost of accuracy. The bright side of visual cues is that they speed up route following when the participants actually follow the guidance. For instance, they could predict turn commands while seeing an intersection several feet ahead, or they would know they have turned to the right direction immediately after seeing an open road in front of them. The negative effect is that over-confidence on visual cues makes the participants underestimate and even ignore the auditory commands when they regard the visual cues as more convincing. This contradiction reveals a problem of interaction: the guidance system expects the participant to travel in a mildly fast and safe manner, typically along a unique path, but the participants sometimes choose the courses that they believe to be more express and relatively safe according to the local visual features.

Even for spatial memory formation, a critical process for active navigation, high visibility did not show a benefit, actually leading to poorer memory compared with low-visibility. The participants' knowledge about the travelled route is gained from spatial updating, which is a multi-sensory [13] and resource demanded process [14] [15]. In a complex and clear visual environment, the off-course visual surroundings might prevent the participants from focusing on the route they took, thus result in shallower processing on the relative positions of the waypoints.

Comparisons on display mode further suggest that the interference from visual cues is a problem of interaction rather than specific display mode. Behaviorally, the vision-related tradeoff has been observed for all the display modes, despite their guidance effects, indicating that the "visual interference" effect might not be eliminated by merely changing display mode of the interface. In addition, the participants had unchanged subjective experience in the high-visibility condition across all the display modes, ignoring their actual effects. This implies that the effect of vision was on the system level. The asymmetry of visual and auditory information might cause confusion. Auditory commands that required participants to go back to the planned route might have appeared unfavorable when they did not want to follow the guidance.

Most of the previous studies focus on improving auditory route guidance performance by using different sound materials and display modes. The current study provides evidence for the possible visual interference on auditory route guidance and raises sensory integration questions within the scope of human-computer interaction. It also sheds light on the design of adaptable auditory route guidance system in the presence of vision. First, since there are usually several possible courses between two waypoints, auditory guidance should provide convincing information, given that the planned route might not appear to be optimal according to the available visual cues. Second, deviation from the planned route in short term is inevitable. It is important for the system to be immune to temporal deviation, and even more to identify voluntary detour and warn the users, thereby eliminating risk-taking behaviors at the early stage.

Acknowledgments. We thank Honeywell International, Inc. for providing the funding. Thanks also to Henry Chen and members of his Honeywell Technology Solutions (Beijing) team for developing the experimental tools.

References

1. Holland, S., Morse, D.R., Gedenryd, H.: AudioGPS: Spatial Audio Navigation with a Minimal Attention Interface. Pers. Ubiquit. Comput. 6, 253–259 (2002)
2. Loomis, J.M., Klatzky, R.L., Marston, J.R.: Assisting Wayfinding in Visually Impaired Travelers. In: Allen, G. (ed.) Applied Spatial Cognition: From Research to Cognitive Technology, pp. 179–202. Routledge, London (2006)
3. Begault, D.R.: 3-D Sound for Virtual Reality and Multimedia, 2nd edn. Morgan Kaufmann, San Francisco (2000)
4. Tran, T.V., Letowski, T., Abouchacra, K.S.: Evaluation of Acoustic Beacon Characteristics for Navigation Tasks. Ergonomics 43, 807–827 (2000)
5. Gaunet, F., Briffault, X.: Exploring the Functional Specifications of a Localized Wayfinding Verbal Aid for Blind Pedestrians: Simple and Structured Urban Areas. Hum-Comput. Interact. 20, 267–314 (2005)
6. Loomis, J.M., Golledge, R.G., Klatzky, R.L.: Navigation System for the Blind: Auditory Display Modes and Guidance. Presence 7, 193–203 (1998)
7. Loomis, J.M., Marston, J.R., Golledge, R.G., Klatzky, R.L.: Personal Guidance System for People with Visual Impairment: A Comparison of Spatial Displays for Route Guidance. J. Vis Impair. Blin. 99, 219–232 (2005)
8. Marston, J.R., Loomis, J.M.: Evaluation of Spatial Displays for Navigation without Sight. ACM Trans. Appl. Percept. 3, 110–124 (2006)
9. Walker, B.N., Lindsay, J.: Navigation Performance with a Virtual Auditory Display: Effects of Beacon Sound, Capture Radius, and Practice. Hum. Factors 48, 265–278 (2006)
10. Giudice, N.A., Bakdash, J.Z., Legge, G.E., Roy, R.: Spatial Learning and Navigation Using a Virtual Verbal Display. ACM T. Appl. Percept. 7(1), 3:1–3:22 (2010)
11. Klatzky, R.L., Lippa, Y., Loomis, J.M., Golledge, R.G.: Encoding, Learning, and Spatial Updating of Multiple Object Locations Specified by 3-D Sound, Spatial Language, and Vision. Exp. Brain Res. 149, 48–61 (2003)
12. Klatzky, R.L., Marston, J.R., Giudice, N.A.: Cognitive Load of Navigating without Vision when Guided by Virtual Sound versus Spatial Language. J Exp. Psychol. Appl. 12, 223–232 (2006)
13. Waller, D., Greenauer, N.: The Role of Body-based Sensory Information in the Acquisition of Enduring Spatial Representations. Psychol. Res. 71, 322–332 (2007)
14. Amorim, M.-A., Glasauer, S., Corpinot, K., Berthoz, A.: Updating an Object's Orientation and Location During Nonvisual Navigation: A Comparison Between Two Processing Modes. Percept. Psychophys. 59, 404–418 (1997)
15. Wang, R.F., Crowell, J.A., Simons, D.J., Irwin, D.E., Kramer, A.F., Ambinder, M.S., Thomas, L.E., Gosney, J.L., Levinthal, B.R., Hsieh, B.B.: Spatial Updating Relies on an Egocentric Representation of Space: Effects of the Number of Objects. Psychon. B. Rev. 13, 281–286 (2007)

Part IV
In-Vehicle Interaction

A Study and Evaluation on Route Guidance of a Car Navigation System Based on Augmented Reality

Kengo Akaho[1], Takashi Nakagawa[1], Yoshihisa Yamaguchi[1], Katsuya Kawai[1], Hirokazu Kato[2], and Shogo Nishida[3]

[1] Advanced Technology R&D Center, Mitsubishi Electric Corporation,
8-1-1, Tsukaguchi-honnmachi, Amagasaki, Hyogo, 661-8661, Japan
{Akaho.Kengo@eb,Nakagawa.Takashi@dy,Yamaguchi.Yoshihisa@dp,
Kawai.Katsuya@dn}.MitsubishiElectric.co.jp
[2] Graduate School of Information Science, Nara Institute of Science and Technology
kato@is.naist.jp
[3] Graduate School of Engineering Science, Osaka University,
1-3, Machikaneyama, Toyonaka, Osaka, 560-8531, Japan
nishida@sys.es.osaka-u.ac.jp

Abstract. We have developed AR-Navi, which is a car navigation system based on augmented reality technology. AR-Navi overlays computer graphics element on live video feed captured by the in-vehicle camera. In this study, we investigated and evaluated information presentation methods in intersection guidance by AR-Navi, from consideration of ease of understanding and also safety. As a result of the evaluations, we confirmed ease of understanding, safety, and the characteristics of AR-Navi in comparison with CG-Navi.

Keywords: Car navigation systems, Augmented reality, Route guidance.

1 Introduction

At present, car navigation systems play a prominent part in the safety and smoothing of road traffic [1]. Many of today's car navigation systems involve route guidance that mainly uses computer graphics (CG). Car navigation systems that use CG are called CG-Navi in this paper. However, CG-Navi cannot completely reproduce the real world that the user sees through the windshield. For that reason, it is difficult for the user to accurately and quickly recognize small differences in location, so the user can get lost or be unable to drive safely because he is peering at the car navigation screen.

Thus, to increase the ease of understanding and safety of car navigation systems, systems have been proposed for providing guidance at intersections, which use an in-vehicle camera to capture live video footage in real time from in front of the vehicle, then overlay CG elements that have been created by using augmented reality technology onto that live video feed.

We are currently developing a car navigation system based on augmented reality technology. The car navigation system we are developing is called AR-Navi [2].

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 357–366, 2011.

To implement this AR-Navi, we must resolve the following two issues:

1. Improved precision of the matching of the live video feed and CG elements
2. Presentation of information about intersection guidance, from consideration of ease of understanding and also safety while driving

To tackle the first issue, we have improved the precision with which the live video feed and CG are overlaid [3].

In this paper, we describe means of tackling the second issue in AR-Navi. We define guidance that enables the user to accurately comprehend easy-to-understand information within a short time and also without error, and investigate a method of presenting AR-Navi information that implements that guidance. In addition, we are studying guidance that does not draw the user's attention away for long, to ensure that the user avoids gazing at the car navigation screen while driving, from consideration of safety. In order to implement those objectives, the following two problems must be resolved:

1. Since AR-Navi guidance is based on various different kinds of information being displayed live, the amount of information on screen tends to become large (*).
   * In this paper, a car navigation screen that causes the user to take a long time between looking at the screen and understanding the information on that screen is defined to have a large amount of information.
2. Since the video feed captured by the in-vehicle camera consists of moving images at 30 fps, the user's eyes will tend to follow the movements in the video when it is displayed on the car navigation screen, and there is a danger that he will gaze fixedly at the screen.

To address the first problem, we have investigated designs for the display of intersection guidance that are easy to understand with a small amount of information. To address the second problem, we have proposed a "best shot" display method as guidance that does not use moving images. In addition, we have implemented a prototype system that incorporates these designs and have conducted driving evaluation experiments on public roads to evaluate the ease of understanding and safety of AR-Navi.

In this paper, we describe the above method and evaluation experiments, then discuss the results of the evaluation experiments.

2 Related Research

We shall now introduce some research into car navigation systems that use augmented reality technology.

Ko et al. have proposed a method of positioning by cross-checking three types of data (road video footage that is input from a CCD camera, the current position of the vehicle, and a road configuration model created from road map data), and have implemented highly precise overlaying of a live video feed and CG[4].

Sawano et al. describes how they showed test subjects some simulated video footage for intersection guidance, which was generated by using augmented reality technology, and guidance displayed by conventional CG-Navi. From comparisons by questionnaire, they found that the display of generated video footage was easier to understand [5]. The research done by Sawano et al. evaluated the ease of understanding of information presented by car navigation systems that used augmented reality technology, which is the same as in our research. They implemented experiments in an environment in which users could gaze at a screen, but there was insufficient investigation in situations in which car navigation system is used while driving.

3 Methods of Presenting Information in Intersection Guidance

3.1 Design of Intersection Guidance Display

To ensure that the user does not gaze fixedly at the screen, the car navigation systems display design must provide display contents that has a reduced amount of information but which enables the user to recognize useful information on the route guidance as simply as possible. We have determined that types of guidance that have any of the following characteristics have a large amount of information:

- Guidance with lots of text and CG
- Guidance using complicated diagrams
- Guidance in which displayed CG elements change dynamically

It is considered that if we could implement a display design which does not have any of the above characteristics and which traces the route by lines and arrows, turn locations could be conveyed intuitively to the user with a small amount of information. Thus we propose the following two types of guidance that focus on the locations at which CG elements are displayed to indicate the route:

1. On-road route display method
2. Overhead arrow display method

1. The distance to Intersection, turning direction
2. Blue symbol indicating the center of Intersection
3. Yellow route or overhead arrow indicating the navigated route
4. Yellow line indicating the navigated route

Fig. 1. On-road route guidance and overhead route guidance

Since the route is displayed on the ground with display design 1, the correspondence with the road is easy to see, but if there are any vehicles ahead, the CG elements are displayed on those vehicles, which give the users an uncomfortable feeling. Since the route is displayed overhead with display design 2, this does not affect any vehicles ahead and there is also little shielding, but it is thought that it will be difficult to see the correspondence with the road quickly, in comparison with display design 1. Each of these designs was implemented as AR-Navi.

An example of the guidance provided by the on-road route display method and the overhead arrow display method are shown in Fig. 1. These examples are just single examples, and it is possible to replace the cone that indicates the intersection center with an arrow and make changes such as displaying side roads.

3.2 Best Shot Display Method

As shown in Section 1, it is dangerous to display moving images on a car navigation screen. Guidelines for the safe handling of image display devices, which have been set out by the Japan Automobile Manufacturers Association, Inc., also state that moving images should not be displayed on such devices [6].

That is why it has been thought that, if the setup is such that the screen is not updated frequently while the user is looking at the car navigation screen, the user will not gaze fixedly at the screen. To that end, we propose a lengthened update period for the moving images, to display single frames at a time. If the moving images are simply displayed as single frames, the generation of adverse conditions such as backlighting could cause the display of an image that is unsuitable for use in the guidance for a fixed time. We therefore propose a best shot display method by which an image that is suitable for the guidance is selected and displayed from the captured moving images.

To implement this method, we study the fundamentals of images that are unsuitable for guidance. When an image captured from an in-vehicle camera is unsuitable, that image is not selected at the "best shot". In this paper, the following conditions cause images to be unsuitable:

1. The image is too dark or bright
2. The user's vehicle is not following the line of the road
3. The user's vehicle and the target intersection of the guidance are too close

Since it is considered that nearer images within the video footage that the user is currently seeing are easier to understand, the best shot display method selects the latest image as the best shot, from among the moving images captured by the in-vehicle camera but excluding unsuitable images.

4 Evaluation Experiments

We created a prototype AR-Navi system by implementing the best shot display method we propose in this paper and intersection guidance display designs, and installed it in a vehicle by mounting it as one function of the car navigation systems. We evaluated the ease of understanding and safety of this AR-Navi, and performed evaluation experiments while driving on public roads to compare conventional CG-Navi and AR-Navi, with the objective of clarifying its characteristics.

4.1 Preliminary Experiments

We performed pilot driving tests on public roads while using AR-Navi, as preliminary experiments. We adjusted the parameters of two items during the runs: the display design and the image update timing spacing of the best shot display method. Since the settings shown in Table 1 were determined to be easy to understand, we implemented the experiments with those settings.

Table 1. Configuration of our system

Setting Items	Configuration
Display Design	On-road route display method
Image update timing spacing	0.75 seconds

4.2 Details of this Experiment

Experimental conditions. The experimental conditions relating to details are listed in Table 2. All test subjects drive on a daily basis and have experience of using car navigation systems.

Table 2. Configuration of experiment

Experimental conditions	Configuration
Test subjects	Three males aged 29 to 32.
Time required experiments	Approximately 50 minutes for each test subject for one display design
Time and weather	Between 11 AM and 4 PM, in both sunny and cloudy weather
Experiment course	Residential areas, ordinary roads with one or two lanes each way
Comparison guidance types	Total of three types: CG-Navi and versions 1 and 2 of AR-Navi

Fig. 2. Examples of car navigation guidance

The display designs used in these experiments are shown in Fig. 2. In these experiments, we used CG-Navi and two versions of AR-Navi with the on-road route display method that was decided upon in the preliminary experiments of Section 4.1. In this paper, these designs are called CG-Navi, AR-Navi Ver. 1, and AR-Navi Ver.

2, respectively. With the display design of AR-Navi Ver. 1, side roads other than the route that the user's vehicle travels on were shown, in addition to the route leading to our company. With the display design of AR-Navi Ver. 2, only the route that the user's vehicle travels on was shown, and side roads were not displayed. AR-Navi Ver. 1 is a display design with a larger amount of data than AR-Navi Ver. 2. To investigate the effects of differences in the amount of information in the display design on safety and ease of understanding, we used two display designs.

Experimental method. The test subjects were given tasks while riding in a vehicle traveling on public roads. Each test subject determined the position and direction to turn from the car navigation guidance, and specified the driving operations. He specified the turn location and direction at the point at which he was confident about the turn location.

Each test subject performed the tasks from the passenger seat, and was instructed to check the forward direction as if intending to drive. In addition, the test subjects were permitted to operate a button that can turn the car navigation screen on or off, to ensure they are not conscious of the screen more than necessary, by turning the display on only when looking at the car navigation screen. The test subjects were also allowed to familiarize themselves with the tasks by performing familiarization runs beforehand.

Evaluation method. We examined the following four items, and also investigated the characteristics of AR-Navi :

1. The results relating to correct/false identification of intersections by the car navigation systems.
2. The visual-contact time ratio for car navigation systems (the total car navigation visual-contact time at a certain intersection, divided by the total time from the start of guidance to that intersection until the vehicle passes the intersection) and the visual-contact time per look
3. Distances from the intersection when the user gives the response when he completely understands the turn location
4. Subjective opinions of test subjects, collected from questionnaires

To study these evaluation items, we recorded each test subject's face, the front of the vehicle, and the car navigation screen as video images, and also measured the times the button was pressed.

4.3 Results and Discussion

First of all, the results relating to correct/false identification of intersections are shown in Table 3 and intersection information generated by false recognitions is shown in Table 4. Since there were few differences in the numbers of false recognitions, we could confirm that there were no major differences between the guidance types in the overall ease of understanding.

We will now discuss the results relating to visual-contact time. The visual-contact time ratio at each intersection when the test subjects were looking at the car navigation screen are shown in Fig. 3 and data relating to single-look visual-contact times is shown in Fig. 4.

Table 3. Numbers of false recognitions of turn locations

	Number of Intersections	Number of falsely recognized intersections
CG-Navi	71	2
AR-Navi Ver. 1	96	3
AR-Navi Ver. 2	91	2

Table 4. Information on intersections that generated false recognitions

Intersection No.	Guidance Type	Characteristics
1	CG-Navi	Intersection of a strange shape
2	CG-Navi	Intersection where the location of the intended turn is a narrow road that is difficult to find.
3	AR-Navi Ver. 1	Intersection with a strange configuration
4	AR-Navi Ver. 1	The road leading to the intersection to turn into is curved.
5	AR-Navi Ver. 1	Within a narrow residential street, with many side roads.
6	AR-Navi Ver. 2	The road to turn into in the guidance was extremely narrow.
7	AR-Navi Ver. 2	Intersection located on uphill slope.

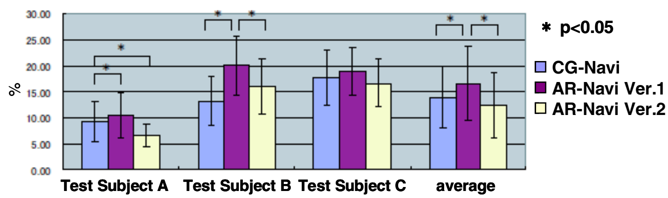


Fig. 3. Average visual-contact time ratios

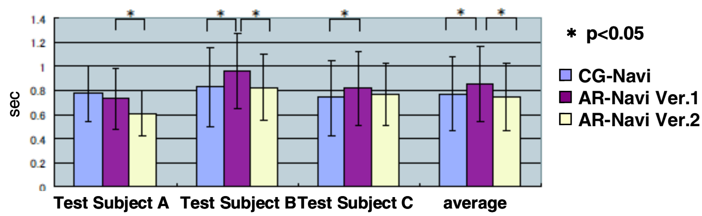


Fig. 4. Average single-look visual-contact times

As the result of performing t-tests at the 5% significance level on averages, we found no significant differences between CG-Navi and AR-Navi Ver. 2 in visual-contact time ratio and single-look visual-contact time when looking at the car navigation screen. The single-look visual-contact time and visual-contact time ratio for AR-Navi Ver. 1 were somewhat higher than for CG-Navi and AR-Navi Ver. 2, so there was a significant difference.

These results are thought to be due to the increased visual-contact time caused by the larger amount of information displayed by AR-Navi Ver. 1, such as side roads.

However, since the single-look visual-contact time average was no more that one second even with AR-Navi Ver. 1, this is thought to be no great problem from the safety viewpoint. In addition, we were able to confirm that there was not great difference in ease of understanding between AR-Navi and CG-Navi.

Results relating to distances from intersection when turn locations are specified are shown in Fig. 5. From the results of t-tests performed between the different types of guidance, significant differences were seen only between AR-Navi Ver. 1 and AR-Navi Ver. 2. These were thought to be due to the lack of information on side roads in AR-Navi Ver. 2. Since there is no information on side roads in AR-Navi Ver. 2, this is thought to be why the subjects were not confident about the turn location until close to the intersection.

As a result, we confirmed that both CG-Navi and AR-Navi enabled confidence in the turn location at about the same distance from the intersection, by displaying side road information. We also confirmed that the effects when there is no side road information have large individual differences.

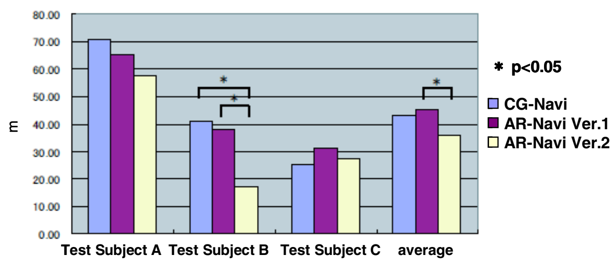


Fig. 5. Distance from intersection at specification

Table 5. Questionnaire results

	Distance from intersection	Time required recognition	Confidence at time of turn	Overall evaluation
CG-Navi	5.00	4.67	5.00	4.67
AR-Navi Ver. 1	4.67	4.33	6.33	5.00
AR-Navi Ver. 2	4.67	4.67	6.00	4.67

The results of questionnaires that the test subjects answered after the evaluation experiments are shown in Table 5. The items covered by the questionnaire were averaged to give questionnaire results for all the test subjects. We asked the test subjects to evaluate each item of the questionnaire in seven steps, with 1 being the worst evaluation and 7 being the best.

From the questionnaire results and open-response questions, we were able to confirm that the overall ease of understanding with AR-Navi was on the same order as that with CG-Navi, and that the guidance was more difficult to understand at some distance from each intersection but because easy to understand closer to the intersection, in comparison with CG-Navi. In addition, the open-response questions showed that all of the test subjects felt there was substantially no difference between AR-Navi Ver. 1 and AR-Navi Ver. 2.

From the results of the evaluation experiments described in this Section, we confirmed that the ease of understanding during rapid visual-contact was about the same for CG-Navi and AR-Navi, when viewed overall. Regarding the safety of AR-Navi, we confirmed that it can be adjusted to about the same as that of CG-Navi, by controlling the amount of information in the display design. We also confirmed that the characteristics of AR-Navi in comparison with CG-Navi are such that the guidance is difficult to understand when far from an intersection but becomes even easier to understand when close to the intersection. In addition, since only a small number of experiments within these experiments related to intersections that generated false recognitions, we performed follow-up experiments relating to the intersections that generated those false recognitions to enable a detailed analysis. We performed follow-up experiments in a desktop environment using the display of a PC. In these follow-up experiments, we have confirmed that AR-Navi is easier than CG-Navi to understand when the road to turn into is narrow and difficult to find, and at intersections where there are many side roads. On the other hand, we also confirmed that AR-Navi is more difficult to understand than CG-Navi with intersections of strange configuration or which have roads proceeding in oblique directions.

5 AR-Navi Improvement Suggestions

From the results of these evaluation experiments, we were able to confirm the characteristics of CG-Navi and AR-Navi, as shown in Table 6.

As a method of resolving the problem that the configuration of an intersection is difficult to understand, we could consider using AR-Navi only when close to the intersection. AR-Navi is easy to understand for giving clear guidance of the turn location close to the intersection, and it also has little misalignment due to vibration of the vehicle and the intersection is not shielded much by buildings or the like.

In addition, the scale of the map displayed on the left side of the screen was made small but in practice this map can be used to help recognize the configuration of the intersection. For that reason, it is thought that this will reduce the problem that the configurations of some intersections are difficult to understand with AR-Navi.

Table 6. Characteristics of each car navigation system type

	When distant from intersection	Directly before intersection	When intersection is shielded	Multiple roads and complicated narrow alleyways
CG-Navi	Easy to understand	Difficult to understand the specific location	Easy to understand	Difficult to determine the specific location
AR-Navi	Difficult to understand	Easy to understand the specific location	Difficult to understand	Easy to identify other roads.

6 Conclusions

In this paper, we investigated and evaluated information presentation methods of car navigation systems in intersection guidance by AR-Navi, from consideration of ease of understanding and also safety.

As a result of the evaluations, we confirmed that ease of understanding AR-Navi is similar between to that of CG-Navi when viewed overall. Furthermore, we confirmed that the safety of AR-Navi would cause no problems if the amount of information displayed is adjusted to be restricted to about the amount of information displayed by AR-Navi Ver. 2. We have also clarified the characteristics of AR-Navi in comparison with CG-Navi.

References

1. Japanese Ministry of Land, Infrastructure, Transport and Tourism ITS HomePage, http://www.mlit.go.jp/road/ITS/j-html/
2. Yamaguchi, Y., Nakagawa, T., Akaho, K., Honda, M., Kato, H., Nishida, S.: AR-Navi: an in-vehicle navigation system using video-based augmented reality technology. In: Proceedings of the 12th International Conference on Human-Computer Interaction, pp. 1139--1147 (2007)
3. Yamaguchi, Y., Nakagawa, T., Akaho, K., Honda, M., Kato, H., Nishida, S.: Evaluation for Adjustment Method of Vehicle's Location by Recognizing Crosswalks. In: Proceedings of the 13th International Conference on Human-Computer Interaction, pp. 648–656 (2009)
4. Hu, Z., Uchida, K.: Solution of Camera Registration Problem via 3D-2D Parameterized Model Matching for On-Road Navigation. International Journal of Image and Graphics 4(1), 3–20 (2004)
5. Sawano, H., Okada, M.: Real-time Video Processing by a Car-mounted Camera and Its Application to Car Navigation Systems by an Augmented Reality-based Display Method. The Journal of the Society for Art and Science 5(2), 57–68 (2006)
6. Japan Automobile Manufacturers Association: Guideline for In-vehicle Display Systems Version 3.0 (2004), http://www.jama.or.jp/safe/guideline/index.html

Evaluation of Collision Avoidance Prototype Head-Up Display Interface for Older Drivers

Vassilis Charissis[1], Stylianos Papanastasiou[2], Lewis Mackenzie[3], and Sachi Arafat[3]

[1] Glasgow Caledonian University
School of Engineering and Computing,
Division of Computing and Creative Technologies,
Glasgow, UK
[2] Chalmers Institute of Technology,
Department of Communication Systems and Information Technology,
Goteborg, Sweden
[3] University of Glasgow, School of Computing, Glasgow, UK
v.charissis@gmail.com, stypap@chalmers.se,
{lewis,sachi}@dcs.gla.ac.uk

Abstract. Spatial and situational awareness could be decreased significantly under low visibility and adverse weather conditions. This could affect exponentially the reactions of the older drivers and increase dramatically their collision probability. To this end we developed a novel Head-Up Display interface that aims to reinstate the drivers' vision which is predominantly hindered under these conditions. In particular the proposed interface entails symbolic representations of the lead vehicles and crucial road information, which effectively enhances driver's vision. The proposed system was evaluated through a comparative study against the typical instrumentation panel. The evaluation results were overall in favour of the prototype interface which improved significantly the reaction times of the older drivers and decreased the collision occurrences.

Keywords: Older drivers, Head-Up Display, HCI, Driving Simulator, low visibility, navigation system, visual interface, collision avoidance.

1 Introduction

Perception and analysis of visual information whilst driving can be a daunting process, particularly for the older drivers, in whom spatial perception and response times tend to be reduced [1]. The latter category represents a rapidly growing segment of the drivers' population requiring additional support in order to counteract these physical and mental age related inefficiencies. Previous studies have demonstrated that the cognitive capabilities and senses deteriorate in geometric progression to age [2]. Interestingly, the produced vehicle interior systems offer a wealth of information and entertainment devices which can additionally hamper even the average drivers' attention.

As such, the over-elaborated, or over-simplified in some rare cases, design of in-vehicle interfaces do not take into consideration the physical and cognitive limitations

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 367–375, 2011.

that older drivers present. In particular, the navigation systems present various road information that older drivers cannot interpret or use in a timely manner [3]. Recently, technological advances produced numerous collision-avoidance systems which have been tentatively and gradually incorporated into contemporary vehicles. Yet the vast majority alerts the driver before an imminent collision and relies heavily on the driver's response time. A small segment of experimental systems even retains the control of the vehicle and regulates the vehicular activities until it is safe to return the control to the driver. Yet both solutions do not provide a viable solution to the older drivers, which either cannot respond in time on the alerts, or they might feel completely intimidated by the temporarily autonomous navigation of their vehicle. Moreover, adverse weather conditions and unexpected traffic congestion can diminish even further older drivers' reaction times, increasing significantly the possibility of collision occurrence and fatalities.

Mindful of the aforementioned observations, our primary objective was to indicate to the user well in advance, potential collisions and subsequently improve the driver's responses without excluding him/her from the actual driving and decision making. To this end, we have developed a full-windshield, Head-Up Display interface, in order to provide the driver with essential, timely presented collision avoidance information [4]. The provided information is superimposed on the windshield through symbolic representation of the forthcoming obstacles and vehicles. The representations are mainly based on the human ability to identify an object based solely on the border-shape of it [1]. Additionally we introduced patterned design colour and size coding in order to intonate the significance of each symbol. Our initial evaluation trials were encouraging, showcasing significantly improved response times and decreased collision occurrences. Yet they provided an insight to previously unforeseen issues that prevented the particular group to capitalize fully the offered information. Through an incremental process of interface design and functionality improvements we managed to incorporate the majority of the older driver's suggestions into the final system. In particular this paper presents the design process of a prototype Head-Up Display (HUD) interface, which aims to improve the driver's spatial awareness and response times under low visibility conditions. The working interface prototype has been further improved for older drivers' use.

Interestingly our previous work towards the development of an all-inclusive users' HUD interface has indicated that such interfaces may enhance significantly the driver's spatial and situational awareness resulting in faster responses and successful collision avoidance [4].

The paper discusses in detail the interface design mantra and presents the results of a large scale comparative evaluation of the HUD system versus the typical dashboard (HDD) information on a group of forty users. Furthermore this work presents an exhaustive analysis of users' age, gender, collision occurrences and the collision speed (m/sec) of each user, which can offer an informative appraisal of the effectiveness of the HUD system through the estimation of collisions per trial, with and without the HUD interface. Concluding, this work describes the HUD interface development towards a user-friendly system which could enhance the response times of the older users.

2 Visual and Cognitive Issues of Older Drivers

The driving process involves multiple senses and cognitive functions and requires precise coupling of mind and body reactions [5].

In particular, visual sensory functions that decline with age are namely; the visual acuity, useful field of vision (UFOV), sensitivity to glare, contrast sensitivity and colour perception [6].

In turn spatial and situational awareness can be detrimentally affected, jeopardising the safety of the older driver [7]. The decline of cognitive and visual sensory abilities challenges the driving performance of older users [8, 9]. More analytically the visual acuity which allows the driver to see objects in detail from a distance decreases gradually after the age of 20 y.o., and declines rapidly after the age of 50 y.o. [10, 11]. Notably this exponential decline of visual acuity is not perceivable by the majority of the elderly driving population [6].

A previous study identified a clear relation between decreased visual acuity and crash rates of older drivers [13]. Older users' vision is further affected by the gradual limitation of the useful field of view, which is the area of the visual field that accommodates information quickly perceivable without any eye or head movements [6]. The significantly reduced visual estate reduces the older drivers' ability to identify objects or movement on the extreme sides of their visual frame, such as vehicles in adjacent lanes [6,7]. The vision sense deterioration of the older users is also affected by the declining ability to distinguish colours [12]. Hence colour perception diminishes significantly up to 50% in the age of 90 y.o [13]. As such the older drivers require clear schematic information which does not rely solely on the colour differentiation between the various states of the information.

Finally the cognitive functions are also affected by age, preventing older drivers to obtain, store and use information [13, 14]. Decreased cognitive ability has an imminent negative impact to the response times (RTs) of older drivers which struggle to make critical decisions under pressure [5, 8].

3 HUD Interface

The aforementioned observations and facts suggested that a novel vehicular navigation system should compensate for the visual impairment of older users and enhance their response time to critical situations without overloading their cognitive functions. As such we strived towards the development of a modular interface which could be accommodated in the direct field of view of the user through the form of a Head-Up Display (HUD). HUD systems have been predominantly used in military aircrafts as they inherently increase screen estate for the real-time presentation of a multitudinous flight and combat information. In tentative steps, HUD has been introduced in the automotive and marine transportation. In particular the early automotive applications offered a new field of experimentation and presentation of additional information collated to the actual driving environment. Concurrent versions investigate the interface conduit which will provide supplementary driving-related information, thus complementing the traditional dashboard, while attracting attention as it is positioned within the driver's immediate field of view.

To this end we have developed a simple interface design which may provide crucial information excluding content that is already presented on the congested dashboard and alert driver for the potential collision risks. The proposed human-computer interface deployed on a full-windshield HUD system, aims to reinstate the three main cognitive factors that affect the older drivers' responses, namely: visual processing speed, ability to divide attention and target clarity [7].

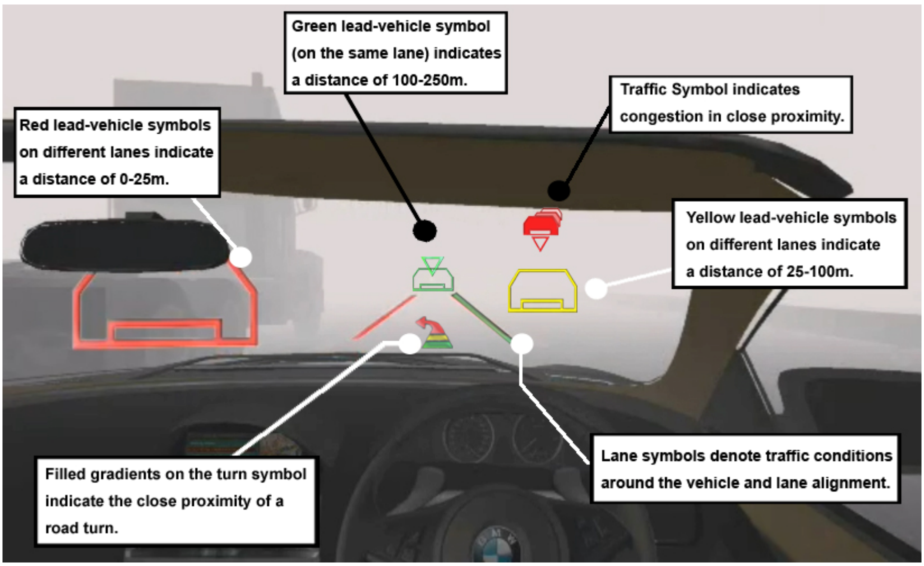


Fig. 1. Screenshot of the actual Driving and brief HUD interface explanatory information

The acquisition and interpretation of a plethora of information available through vehicular sensors, enables the HUD interface to convey this information to the user in an uncluttered and simplified visual output. As such the system could enhance the perception of the vehicle's surrounding space and improve the driver's response times, particularly under low visibility conditions. Our initial experimentations and the subsequent evaluations, informed further our interface design and functionality regarding the fine tuning of the quantity, quality, timing and presentation of the visual cues.

In tandem, the system offers symbolic representation of the vehicles and surroundings of a motorway environment, such as lead vehicles, lane recognition, sharp turn warning and traffic alert. These symbols trace the actual vehicles, lane positioning and curvature and embed themselves to the environment through the Head-Up Display projection system as illustrated in Figure 1. Their size and colour-coding alters respectively with the distance and time to collision (TTS) parameters provided by the vehicular ad-hoc network system (VANETS) that gathers this surrounding information.

4 Modelling and Simulation

In order to evaluate the improved prototype interface, a custom driving simulator was developed to investigate the efficiency of a full-windshield Head-Up Display (HUD) interface for the elder drivers' in low visibility and traffic congestion conditions. The driving-simulator (OSDS) was based on an open source code and off-the-shelf hardware components [4].

The numerical information acquired during the experiment traced primarily the actual response times (RT) and headway (HW) benefits derived with and without the HUD usage. However being aware of the plethora and complexity of factors involved in an imminent collision such as visibility, psychological refractory period, spatial awareness, situational awareness and stimulus-response compatibility amongst others, we employed various methods of recording the human responses such as heart-rate, eye-tracking, body posture and facial expressions.

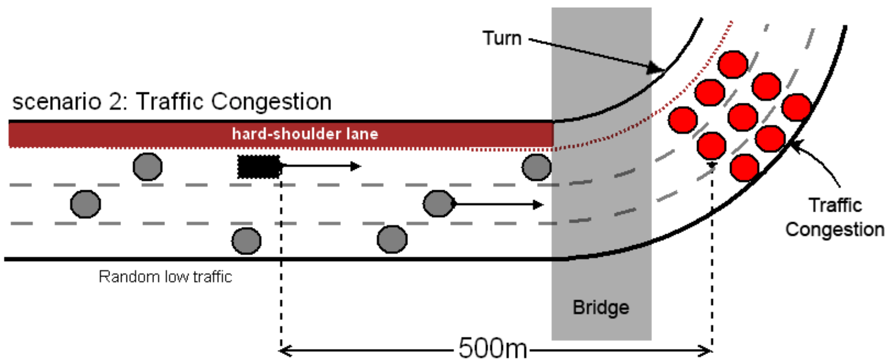


Fig. 2. Traffic congestion and sharp turn situations (second scenario)

The simulation scenario entails low visibility, sharp turn and traffic congestion practically invisible positioned behind a bridge maximizing the effect of surprise as depicted in Figure 2. In addition the agent vehicles of the normal flow and the traffic congestion were infused with a degree of artificial intelligence in order to increase the realism of the simulation scenarios [15]. Their individual driving patterns embedded on the agent vehicles were derived from real-life observations and driver behaviours defined by proprietary literature. The different accident scenarios were kindly provided by the Strathclyde Police Department in Glasgow.

Notably in this scenario the drivers performed adequately in comparison to other simulated scenarios. Yet the collisions occurrences were still at alarming levels.

In accordance to our previous evaluation methods the second scenario trials attracted 40 volunteer drivers with valid driving licence and aging between 20 and 75 years old. The drivers were randomly selected in order to represent the largest possible part of driving population spanning across a large variation of driving experience, profession, gender and age.

The primary aim of the experiment's design was to evaluate the benefits, if any, of using the HUD during driving under low visibility conditions in a motorway environment with particular interest to improve significantly the response times of the elder users.

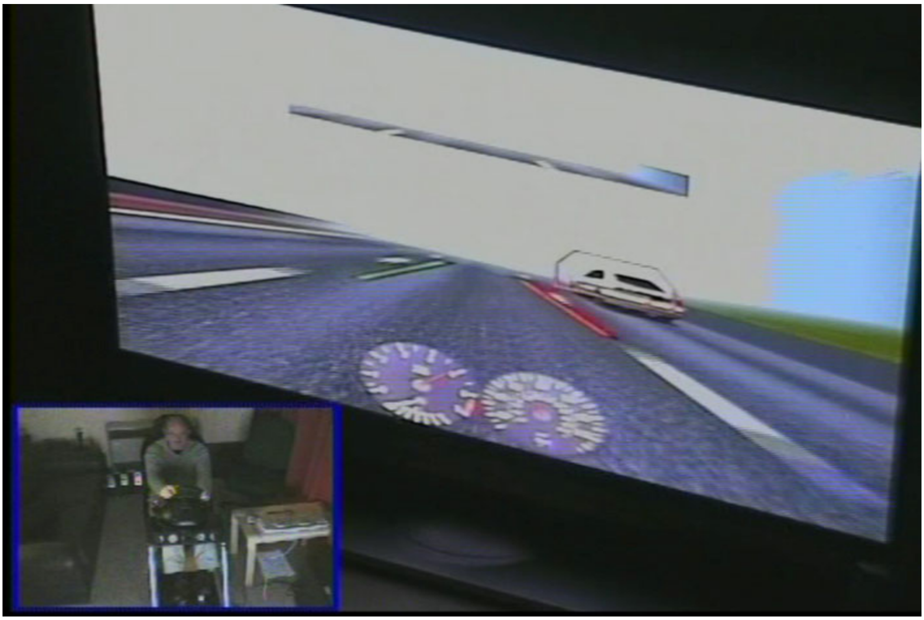


Fig. 3. Screenshot of the user video footage and the driving simulator in action

5 Results and Discussion

The analysis of the numerical data offered a precise view of the drivers' reactions, benefits and drawbacks of the proposed HUD interface in contrast to the existing HDD system. Yet further observation of the video footage and the subjective feedback verifies the numeric information and highlights the overall user experience during the different stages of the experiment. In particular, the video-captures presented distinguishable body-postures during the simulation of the accident scenarios with and without the HUD system.

Further observations highlighted that the drivers without the use of HUD support, tended to keep their elbows closer to the body, hold tight the steering wheel and in some cases, leaning forward on the driving seat. Their anxiety was also evident through their facial expressions. During the subjective feedback and interview they noted that the low-visibility due to the simulated fog produced a very difficult and dangerous driving environment which was very uncomfortable and distressing.

Interestingly, the older drivers recognised the symbols' functionalities and familiarised themselves with the HUD interface within minutes. The users commented positively on the simplicity of the system and the unusual comfort of not needing prolix manual instructions, typically accompanying the majority of navigation systems and other infotainment devices. This was a common observation made by all the forty drivers that participated in the experiment. Furthermore the users commented positively on the full-windshield projection, which immersed the highlighted symbols with the actual scenery, enabling the older users to avoid imminent collisions.

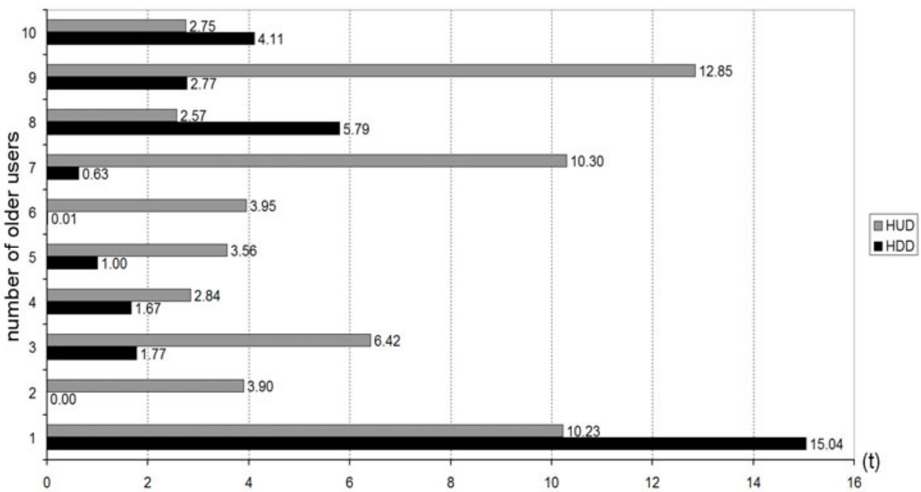


Fig. 4. Comparison of collision speeds (a) with HUD and (b) with HDD

The subjective feedback was also supported form the numerical findings. Focusing particularly on the older drivers group, the effect of the HUD usage became immediately visible as the simulation recorded significantly decreased collision occurrences.

An analysis of their driving behaviour through the numerical data revealed that the older users managed to maintain a sufficient time-to-collision (TTC) window with the lead vehicle/obstacle with the use of the HUD interface. In contrast the use of HDD did not offer sufficient information which resulted in a significantly minute TTCs and increased collisions as presented in Figure 4.

Notably users 4, 7 and 10 although maintained their distance from the lead vehicles they lost the control of their vehicle in the road-curve negotiation, although they managed to maintain their distance from the lead vehicles and the traffic congestion. As such 50% of the older drivers that participated in this trial collided either with the vehicles forming the traffic bottleneck or with the difficult hairpin type of road curve.

In contrast none of the drivers collided with these simulation elements when the HUD interface was activated. Another interesting finding was the follow-up speed of the lead vehicles that the older users maintained which was constantly higher with the use of the proposed interface as depicted in Figure 5, without though exceeding the motorway limits.

This can be attributed to the fact that the highlighted information by the HUD interface offered a significant advantage to the drivers and particularly to the older population by enhancing effectively their vision. In turn the users could trace the traffic and the indistinguishable objects more than 250m ahead even in low visibility

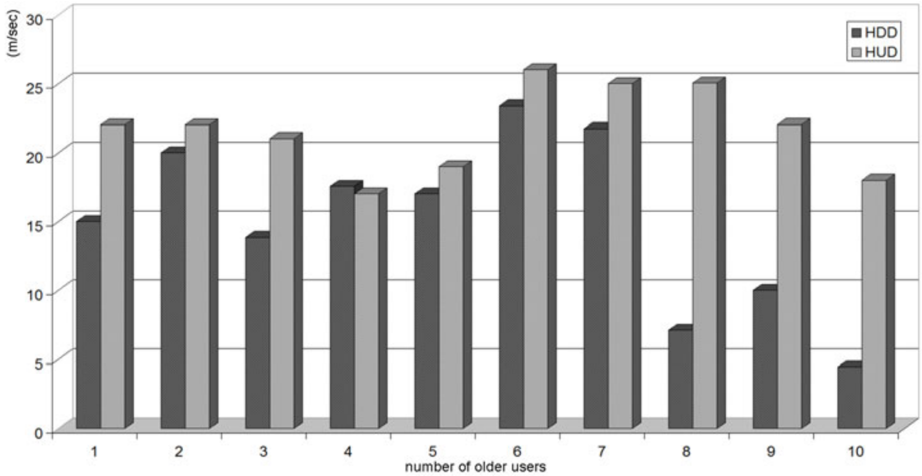


Fig. 5. Comparison of follow-up speeds (a) with HDD and (b) with HUD

conditions. As a result the drivers could produce the typical motorway speeds, yet remaining calm and in control of their vehicle.

6 Conclusions

This paper offered a succinct presentation of the contemporary issues hindering the older driving population. In turn, we presented our proposed guidance system which entails a prototype HUD interface deployed in full-windshield. Adhering to the typical spatial and situational awareness issues of the older drivers the system provides the users with simplified visual information, yet crucial for the effective collision avoidance. The evaluation of the system against the contemporary instrumentation panels, suggests that the proposed HUD improves significantly the users' reactions and effectively decreases the collision occurrences. However the system does not currently offer any customisation attributes. As the majority of the drivers increased substantially their speed the system should be enabled in the future to advise the user for the optimal driving speed in order to avoid abuse of the system's benefits.

Finally our future plan is the incorporation of auditory and haptic cues which will complement the visual HUD warnings. Furthermore we envisage the development of an automated calibration system which will optimise the visual HUD inputs according to the visual characteristics of each user.

References

1. Hoffman, D.D., Richards, W.A.: Parts of Recognition. In: Pinker, S. (ed.) Visual Cognition, pp. 65–96. MIT Press, Cambridge (1985)
2. Rakotonirainy, A., Steinhardt, D.: In-vehicle technology functional requirements for older drivers. In: Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Essen, Germany, pp. 27–33 (2009) ISBN:978-1-60558-571-0

3. Itoh, N., Yamashita, A., Kawakami, M.: Effects of car-navigation display positioning on older drivers' visual search, in Assessment and Promotion of Work Ability, Health and Well-being of Ageing Workers. Journal of International Congress Series 1280, 184–189 (2005)
4. Charissis, V., Vlachos, G., Chan, W., Arafat, S.: Early Notification Warning System for Prototype Head-Up Display: Development and Evaluation of Traffic Congestion and Sharp Turn Warnings. In: Proceedings of the Society of Automotive Engineers (SAE) World Congress, Detroit, Michigan, USA, April, 12-15 (2010)
5. Musselwhite, C.B.A., Haddad, H.: Exploring older drivers' perceptions of driving. European Journal of Ageing 7(3), 181–188 (2010)
6. European Road Safety Observatory (2006) Older Drivers, `http://www.dacota-project.eu/Links/erso/knowledge/Fixed/07_old/olderdrivers.pdf` (retrieved, November 10, 2010)
7. Ball, K.: Attentional Problems and Older Drivers. Alzheimer Disease and Associated Disorders 11(suppl. 1), 42–47 (1997)
8. Clarke, D.D., Ward, P., Bartle, C., Truman, W.: Older drivers' road traffic crashes in the UK. Accident Analysis Prevention 42(4), 1018–1024 (2010)
9. Ryu, J., Song, G., Lee, S., Cho, Y., Kyung, G., Kim, H., Baek, K.: Improving in-Vehicle Display and Control Design for Older Drivers. In: Karsh, B.-T. (ed.) EHAWC 2009. LNCS, vol. 5624, pp. 170–176. Springer, Heidelberg (2009)
10. Meyer, J.: In-Vehicle Telematics Systems and the Older Driver. The Silver Market Phenomenon 2008, IV, pp. 381–390 (2008)
11. Owsley, C., McGwin Jr., G.: Vision and Driving. Vision Res. 50(23), 2348–2361 (2010)
12. Eby, D.W., Trombley, D., Molnar, L.J., Shope, J.T.: The Assessment of older driver's capabilities: A review of the literature (Report No. UMTRI-98-24). University of Michigan Transportation Research Institute, Ann Arbor, MI (1998)
13. Bayam, E., Liebowitz, J., Agresti, W.: Older drivers and accidents: A meta and data mining application on traffic accident data. Expert Systems with Applications 29, 598–629 (2005)
14. Caird, J.K., Chisholm, S.L., Lockhart, J.: Do in-vehicle advanced signs enhance older and younger drivers' intersection performance? Driving simulation and eye movement results. International Journal of Human-Computer Studies 66(3), 132–144 (2008)
15. Charissis, V., Papanasasiou, S.: Human-machine collaboration through vehicle head up display Cognition. Technology & Work 12(1), 41–50 (2008)

The H-Metaphor as an Example for Cooperative Vehicle Driving

Daniel Damböck, Martin Kienle, Klaus Bengler, and Heiner Bubb

Lehrstuhl für Ergonomie, Technische Universität München, Boltzmannstr. 15, 85747 Garching bei München, Germany
{damboeck,kienle,bungler,bubb}@lfe.mw.tum.de

Abstract. For quite a while the automotive industry has been working on assistance systems to improve safety and comfort of today's vehicles. In the course of this development combined with increasingly capable sensors, assistance systems have become more and more powerful. This whole development enlarges the role of the human, beginning from the actual driver of the car up to a supervisor of the automation state. On the one hand this leads to a relief in the drivers workload. On the other hand effects like out-of-the-loop and associated with that a loss of situation awareness can appear. Trying to solve this clash of objectives, the project "H-Mode" follows an idea of vehicle driving where the automation is capable of driving almost autonomous, but the driver is still kept active and in the loop by cooperating with the automation-system. The article describes the idea of cooperative driving and especially the H-Metaphor. Furthermore an example is given how this concept is used in the development of assistance and automation systems.

Keywords: highly automated driving, driver assistance, shared control, haptic feedba0063k, cooperative control, side stick.

1 Introduction

Functional progress in the area of advanced driver assistance and active safety systems (ADAS) enables higher degrees of automation in future cars. A comparable development has been seen in aviation and maritime scenarios. Meanwhile ADAS like adaptive cruise control, lane keeping assistant systems and collision avoidance are widespread and assist the driver partially and in given situations. But for a long time being – like in aviation – the human operator, i.e. the driver, will have to be kept in the loop due to the Vienna Convention. On the other hand the technical potentials for safety and comfort shall be gained. Therefore it is an aim to investigate ergonomic human machine interfaces that avoid typical automation effects but also increase the comfort for the user of a highly automated vehicle.

One aspect is to ensure that the interaction concept fulfills the criterion to be compatible with the driver's expectations and with the environment in which the driving task has to be fulfilled in cooperation with a very powerful cognitive enabled machine. Moreover it is necessary that as much information as possible is transferred between the person and the machine considering actual and future status of the driving task and future intentions on further driving maneuvers.

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 376–385, 2011.

According to this, the challenge for highly automated vehicles is to reduce a relatively high complexity of the automation into a manageable complexity for the human being. Aviation can only be a limited role model here: In most aircrafts, two well-trained pilots keep the system safe, a "luxury" in redundancy that is usually not available in ground vehicles. New concepts for an intuitive approach to automation that everybody can operate without extensive training have to be developed and tested.

One potential technique to increase intuitiveness is the use of design metaphors. In the computer domain, the desktop metaphor took a natural desktop as an inspiration for the organization of a PC user interface with folders, trash cans etc. For intelligent vehicles, the H-Metaphor takes the example of the rider-horse relationship to describe a cooperative interaction between a highly automated vehicle and a driver (H-Mode). Initially developed for air vehicles [1] [2] it is now systematically applied to cars and trucks [3].

The horse as a role model is suited with sufficient intelligence that can be used to allude the rider to changes according to the environmental setting, to influence his behavior, to widely take over control in non-critical situations and maybe even react on its own in critical situations. Transferred to the vehicle, an intelligent automation can act likewise. By becoming increasingly capable due to technical improvements automation-systems are getting more and more "intelligent" up to the point where they can actively affect the driver's behavior and release him in situations where he is overstrained.

Therefore both the driver and the automation need to interact with each other on a cooperative basis [4]. To do this an interaction concept is needed where both the driver and the automation can communicate wishes and pieces of advice to the respective partner and negotiate a common course of action. An essential aspect thereby is that the driver and the automation are simultaneously involved in the driving task, acting parallel to each other (Fig. 1).

Both perceive the environment separately, generate an intention based on this perception and try to put this intention into practice by affecting the vehicle, the driver or accordingly the automation. Following the H-Mode interaction paradigm, this communication and negotiation is primarily carried out via the manual haptic channel using active control elements. In addition to this the haptic interaction is supplemented with acoustic and visual information.

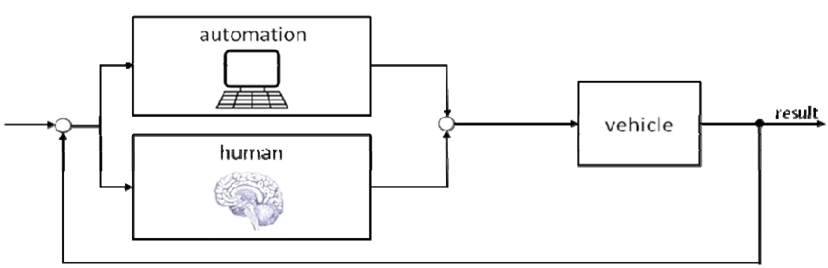


Fig. 1. Schematic drawing of the parallel-simultaneous interaction concept. Driver and Automation act parallel to each other and interact with each other via a summation point.

2 Active Control Elements for Cooperative Interaction

Active control elements in general provide a way to benefit from haptic feedback [5]. Forces which can be generated by the integrated actuators can be used to transmit vital information to the operator. Therefore the mechanical connection between the machine and the operator can be separated and replaced by an electronic one. On the one hand this decoupling makes it possible to completely redesign the interface. On the other hand the separation from the mechanical feedback may aggravate the user's ability to operate the system. The loss of information flow results from the fact that the operator can only feel the dynamics of the control element, but not the dynamic of the controlled system itself. Therefore the user has to estimate the system's behavior in order to keep the system within safety limits [6]. For technical purposes, two concepts of active operating elements must be distinguished: force and position reflective elements.

In the following these drafts are exemplified by driving a side stick based vehicle. A side stick as control element has been chosen due to the fact that from an ergonomic point of view the dimensionality of a control element should correspond with the dimensionality of the task. According to this a two-dimensional side stick should be suitable for longitudinal and lateral guidance in car driving.

For driving the vehicle the operator creates forces on the stick. The underlying spring characteristic of the force reflective operating element (Fig. 2) determines its movement as a function of the load injected by the operator. Through the stick position the user adjusts the set point settings of the vehicle.

Consequently the dynamics of the stick is autonomous and does not allow conclusions about the vehicle's state. This means for example in lateral direction that the driver manipulates the steering angle but has no knowledge about its actual state. He can only estimate the wheel position through the sensed accelerations.

On the contrary position reflective elements (Fig. 3) use the applied forces to generate the set point settings [7]. As opposed to the spring centered stick, where the position results from the balance of forces, the position reflective control element stays fixed for the operator and is only moved by the controlled system. More precisely the forces applied by the driver are measured and transformed into control inputs. The feedback information is returned as position of the element which thereby represents the actual state while its movement represents the dynamic of the system itself. Consequently the operator senses the behavior of the system/vehicle.

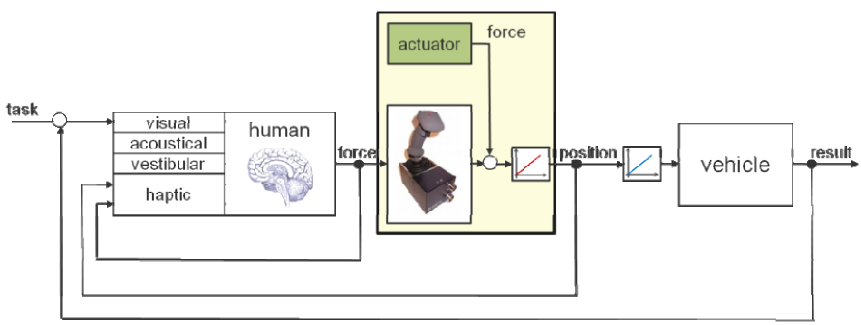


Fig. 2. Force reflective control element

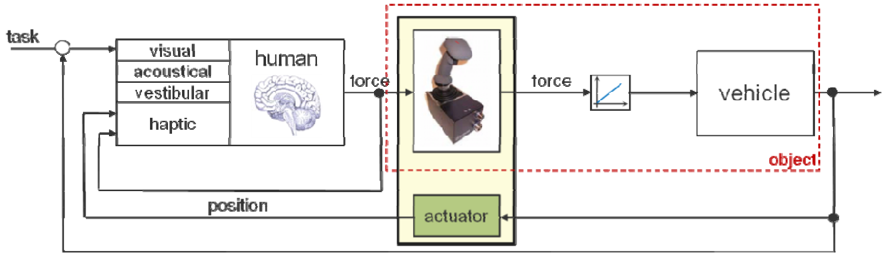


Fig. 3. Position reflective control element

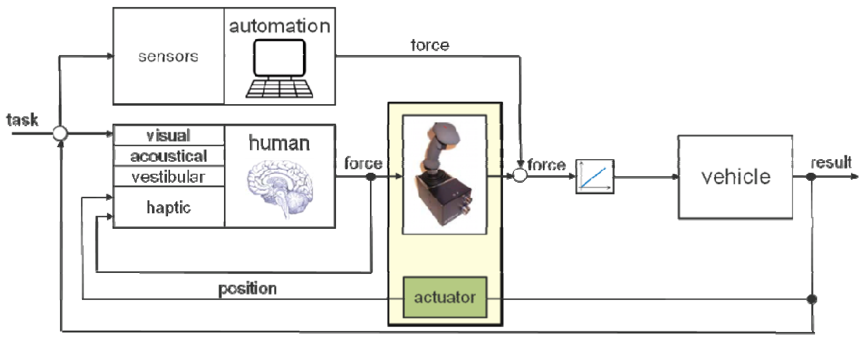


Fig. 4. Combination of position-feedback of the vehicle dynamics and force-feedback of automation recommendations

This configuration works similar to the direct interaction with objects and direct manipulation. By raising forces the user manipulates the item which responds with movement. As a result, position reflective elements seem not only suitable for compensating the decrease of information flow but also enable a specific feedback of essential information that supports the operator. Applied to the task vehicle driving a significant advancement in driving performance can be verified [5].

One of the essential features of the H-Mode is a bi-directional haptic-multimodal coupling with continuous and/or discrete communication between driver and automation. This communication has to be established in addition to the feedback of the system's dynamics. Thereby the driver will be kept in the loop. Status and recommendations of the automation will be communicated. This is not only a feature driving on a high level of automation but is also reasonable on low levels of automation. Fig. 4 shows the combination of the position-feedback principle with a feedback of automation recommendations [8]. In the same way the driver applies forces to the control element, the automation can influence the vehicle's behavior by generating virtual forces. Both inputs are merged in a summation point and the resulting force is used to generate the set point settings for the vehicle. Via the feedback-loop the driver can haptically sense the generated forces of the automation on the control element i.e. he can sense its recommendations.

Experiments show that the combination of both feedback principles – position feedback of the vehicles dynamics and force feedback of automation recommendations – grants a further benefit to the performance of the driving task.

3 Degrees of Automation

One important component of the H-metaphor is the possibility of the user to be able to change between different levels of automation from high automation to low/no automation covering the whole range of different levels of automation (Fig. 5).

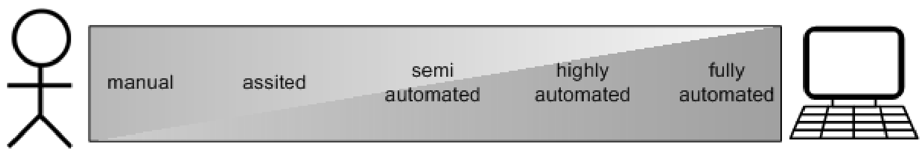


Fig. 5. Involvement in system control as a spectrum between fully manual and fully automated [9].

Therefore the interaction concept has to provide a possibility to switch between the degrees of automation. With an increasing degree of automation the system gains more and more influence on the driving task, thus meaning that the driver hands over authority to the automation. Subsequently, the driver gets more authority and a more direct control of the vehicle with a decreasing degree of automation. In terms of realization this corresponds with a variable weighting of the input – driver and automation – at the summation point (Fig. 2).

In the original domain, this allocation of authority is dependent on how the horse reins are used. If the rider wants to have more impact on the horse's actions, he tightens the reins and thereby vastly takes over control. In terms of vehicle driving this "Tight-Rein-Mode" corresponds with a low level of automation. For a more indirect control, respectively a transfer of authority to the horse the rider eases the reins. This "Loose-Rein-Mode" equals a high level of automation.

Based on this interaction paradigm a prototypical system has been implemented in which the grip force is used as an indicator for how much the driver wants to be actively involved in the driving task [10]. Therefore the stick is equipped with a sensor system to measure the actual grip force of the driver. Corresponding to the tight reins of the rider the driver increases the grip force applied to the control element and thereby raises his own authority in the driving task. If the driver wants to grant the automation higher latitude, he reduces the applied grip force and therefore transfers authority to the automation. This relation can be realized not only for two states but it could have the potential to be scaled on a continuum. The following chapter shows an usability-study about grip-force-measurement with three changeable degrees of automation.

4 Grip Force as Indicator for Driver Involvement

This experiment investigates whether grip force can be used as a nonverbal parameter to differentiate between different levels of driver involvement and serve as a switch for the level of automation. Based on the preliminary results, this experiment is performed with a position reflective side stick with steering angle feedback in lateral and speed feedback in longitudinal direction.

Fig. 6. Usability laboratory at the department of Ergonomics

4.1 Method

Experimental Setting: The experiment takes place in the H-Mode usability laboratory at the Institute for Ergonomics, Technische Universität München (Fig. 6). The laboratory consists of a highly variable mockup with a single projection screen. The driving simulation software SILAB TM directly receives all commands from the control elements, provides the vehicle and driving dynamics simulation and generates the information for haptic feedback at the stick. Data sampling and logging facilities are also provided by the SILAB software.

Two different concepts for the change of automation degree are compared. They differ in the method to initiate the transition of the automation degree. Whereas in the first version the shift between tight rein and loose rein can be accomplished by pressing the associated button. The second prototype measures the operator's grip force applied to the control element and changes the degree of automation accordingly. In order to measure the grip force the conventional side stick grip is substituted for a rudimentary grip with force-sensing resistors (Fig. 7, left).

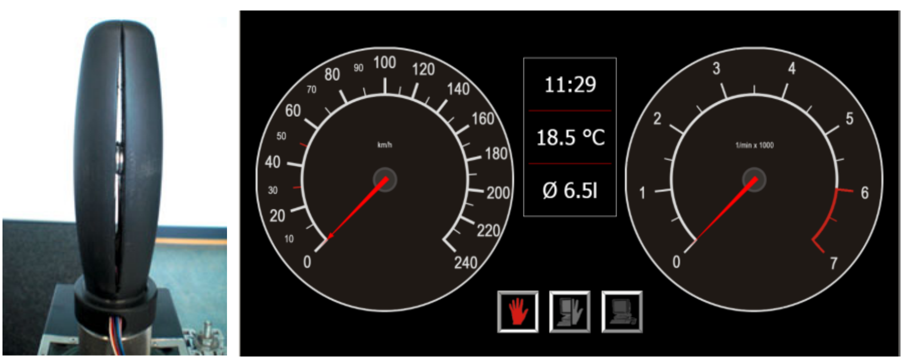


Fig. 7. Side stick grip with FSR sensors (left). Touch screen dashboard with buttons for completely manual, semi automated and highly automated. (right).

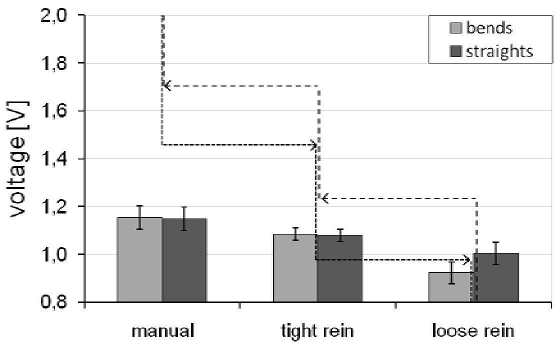


Fig. 8. Bar chart: measured grip force from preliminary tests (mean ± standard error, N=24), lines: realized borders for switches between automation degrees.

The voltage gradient at each of its four sensors is proportional to the pressure applied. The average of all four sensors is used as reference value for the automation degree. A high voltage thereby is equal to a high pressure, i.e. a low automation degree and vice versa. This means that a high grip force reduces the influence of the assistance while loosely holding the side stick grants the automation the possibility to take major control over the vehicle. For alternative experimental conditions the virtual dashboard which is based on a touch screen, was extended with three buttons providing an interface for initiating transition in the automation spectrum (Fig. 7, right).

The automation thereby consists of three levels: completely manual, semi automated and highly automated. The difference between the two automated levels is mainly based on the forces the automation may apply to the control element. Based on a preliminary test a system with grip force induced automation shift was implemented. In order to alleviate the intentional shift of the automation degree the force levels have been elevated . To prevent a continuing alternation of the automation state at the border between two defined states a characteristic hysteresis curve has been added (Fig. 8).

Participants: 16 subjects (13 male, 3 female) with an average age of 29.5 years (standard deviation: 5.3 years) participate in the experiment. All subjects have a driving experience of at least 5.000 km/year and normal or corrected-to-normal visual acuity.

Experimental design: The experiment is set up using a within-subjects design with the switching method as independent variable and different items of subjective acceptance as dependent variables. For the assessment a semantic differential consisting of various Likert-type bipolar rating scales is used for the assessment of subjective acceptance covering a range from -3 (strongly disagree) to +3 (strongly agree). As this experiment is only thought to show the potential and subjective usability of the grip force measurement only subjective data are statistically assessed. Therefore the different items are compared pairwise with a t-test. The level of statistical significance is set at $\alpha = .05$.

Experimental procedure: As most of the subjects have never participated in a driving simulator study before and as none of them has had any experience driving with a side stick, the experiment starts with a training session of about 20 minutes in which the subjects can get used to the simulated environment and to the unfamiliar control concept (including side stick and changeable degrees of automation). After this training run, two trial runs are performed in permuted order during where subjects have to change the degrees of automation either by pressing buttons or by varying the grip force (depending on the run). The subjects are instructed to change to the automation mode according to the mode presented beside the road on traffic signs. Both runs are performed on the same test track (30 kilometer road, partially winding). Every run is accompanied by a questionnaire regarding the subjective acceptance.

4.2 Results

The evaluation of the subjective questionnaire regarding 5 different items of subjective acceptance is presented in Fig. 9. The results of the items comfortable ($t(15) = -0.436$; $p = .669$) and simple ($t(15) = 0.000$; $p = 1.000$) show no significant effect. Regarding the items feasible ($t(15) = 2.298$; $p = .036$), attractive ($t(15) = 4.162$; $p = .001$) and even sportive ($t(15) = 1.986$; $p = .047$) the grip force measurement is rated significantly better.

Summarized it can be said that the grip force measurement is not rated significantly worse in any of the items but is rated better in three of five. This result seems promising regarding further investigations on this topic.

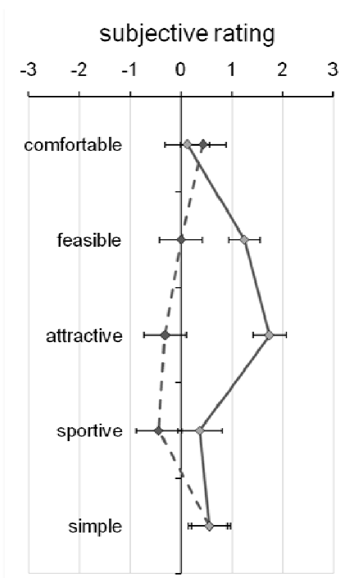


Fig. 9. Subjective rating of the different transition modes (mean ± standard error, N=16). Dashed line: buttons, continuous line: grip force

5 Conclusion

In the long run completely autonomous cars are not very likely to be on the market due to technical restrictions (e.g. sensor range) as well as the Vienna Convention which says that the driver must be kept in the loop. Nevertheless the increasing number and quality of driver assistance systems enables the feasibility of a highly cooperative vehicle. One potential implementation of this idea follows the H-Mode interaction paradigm as a parallel-simultaneous interaction concept. Taking this parallel interaction as a basis, the question arises, whether conventional control elements are conducive and if there might be any considerable alternatives, especially regarding the human-automation-communication.

The experiment shows how design metaphors can be used to generate and develop new and unconventional interaction ideas. Its focus is a usability test to assess the idea whether grip force can be used as a parameter to differentiate between different levels of driver involvement and therefore serve as a switch for the level of automation. This method is compared to a touch screen with three buttons providing an interface for initiating transitions in the automation spectrum. The comparison shows positive results throughout all attributes. Therefore the initial idea seems to have potential and will be further pursued. However many questions remain especially how does the grip force work in situations of surprise, stress or shock. The apprehension arises that in such situations humans show a tendency to hold onto the control elements and thereby increase their grip force. If this were the case, the relation “high grip force – low assistance” would be extreme counterproductive, as the driver would especially need more assistance in these situations. Nevertheless should be further investigated if the potential of grip force measurement can be used as a nonverbal additional information in combination with other channels (multimodal fusion).

More experiments regarding different control elements besides the side stick and the steering wheel will also be conducted. Conventional elements of driving cannot be substituted overnight by side sticks. Thus other possibilities have to be explored. Alternatives suitable for migration might already exist in other domains but have to be found.

References

1. Flemisch, F.O., Adams, C.A., Conway, S.R., Goodrich, K.H., Palmer, M.T., Schutte, P.C.: The H-Metaphor as a Guideline for Vehicle Automation and Interaction (No. NASA/TM-2003-212672). Hampton: NASA. Langley Research Center (2003)
2. Goodrich, K.H., Schutte, P.C.: A Design and Interaction Concept for Aircraft with variable Autonomy: Application of the H-Mode. In: 25th Digital Avionics Systems Conference (2006)
3. Flemisch, F.O., Schieben, A.: Highly automated vehicles for intelligent transport: Validation of preliminary design of HAVEit systems by simulator tests. Deliverable to the EU-commission D33 3 (2010)
4. Flemisch, F.O., Kelsch, J., Löper, C., Schieben, A., Schindler, J., Heesen, M.: Cooperative control and active Interfaces for vehicle assistance and automation. FISITA World Congress (2008)

5. Kienle, M., Damböck, D., Kelsch, J., Flemisch, F.O., Bengler, K.: Towards an H-mode for highly automated vehicles driving with side sticks. Automotive User Interfaces 2009, Essen, Germany (2009)
6. Schieben, A., Damböck, D., Kelsch, J., Rausch, H., Flemisch, F.: Haptisches Feedback im Spektrum von Fahrerassistenz und Automation. 3. Tagung Aktive Sicherheit durch Fahrerassistenz (2008)
7. Damböck, D., Kienle, M., Flemisch, F.O., Kelsch, J., Heesen, M., Schieben, A., Bengler, K.: Vom Assistierten zum Hochautomatisiertem Fahren - Zwischenbericht aus den Projekten DFG-H-Mode und EU-HAVEit. VDI-Congress Fahrer im 21. Jahrhundert (2009)
8. Damböck, D., Kienle, M., Bengler, K.: Bedienelemente hochautomatisierter Fahrzeuge – Studie zum haptischen Kanal als Kommunikationsmedium. Frühjahrskongress der Gesellschaft für Arbeitswissenschaft (2010)
9. Flemisch, F.O., Kelsch, J., Löper, C., Schieben, A., Schindler, J.: Automation Spectrum, inner/outer compatibility and other potentially useful human factors concepts for assistance and automation. In: de Waard, D., Flemisch, F.O., Lorenz, B., Oberheid, H., Brookhuis, K. (eds.) (Hrsg.) Human Factors for Assistance and Automation (2008)
10. Damböck, D., Kienle, M., Bengler, K.: Die Zügel fest in der Hand halten – Automationsgradumschaltung durch Griffkraftmessung. VDI-Congress, Useware (2010)

Factors for Representing In-Vehicle Roominess

Wonil Hwang[1], Nam-Hyo Kim[2], Hyeong-Joon Ahn[3], and Hee-Seok Jung[4]

[1] Department of Industrial & Information Systems Engineering, Soongsil University, Seoul, Korea
[2] School of Architecture, Soongsil University, Seoul, Korea
[3] Department of Mechanical Engineering, Soongsil University, Seoul, Korea
[4] Research & Development Division, Hyundai & Kia Corporate, Korea
{wonil,pratt95,ahj123}@ssu.ac.kr,
ftj05@hyundai.com

Abstract. Car drivers or passengers tend to perceive the interior space of motor vehicles as a psychological space rather than a physical space. Even though cars have the same or similar volumes of the interior space, car users may perceive different in-vehicle roominess according to the characteristics of the interior space of the cars. In this study, we aim to investigate factors that represent in-vehicle roominess as psychological dimensions. 7 experts participated in collecting 105 psychological expressions for in-vehicle roominess, and the experiments were conducted with 15 participants and 7 diverse motor vehicles. Through factor analysis and multiple linear regression, we found 10 factors and 3 most influential factors in the in-vehicle roominess: namely, 'space completeness', 'narrowness' and 'dullness'. We anticipate utilizing these factors for designing the interior space of motor vehicles in terms of psychological dimensions.

Keywords: Interior Roominess of Vehicles, Psychological Dimensions, In-vehicle Roominess Factors.

1 Introduction

In everyday life, people spend lots of time in motor vehicles. While riding the motor vehicles, car drivers and their companies perceive and feel the interior space of the motor vehicles and get a variety of subjective evaluations based on the space sense, including narrow/wide, small/big and so forth. So far, the interior space of the motor vehicles has been treated as a physical space, and studied using the measures based on mechanical dimensions. However, car drivers or users tend to perceive the interior space of the motor vehicles as a psychological space rather than a physical space, and thus, they express their perceptions and feelings for it in various ways. Even though several cars have the same or similar volumes of the interior space of the motor vehicles in physical or mechanical dimensions, car users may perceive different, so-called, 'in-vehicle roominess', according to the characteristics of the interior space of the cars in psychological dimensions.

Until now, regarding the design of interior space of motor vehicles and driver's or user's perception, there have been three research streams. The first research stream

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 386–390, 2011.

focused on how efficiently the interior spaces of cars were designed and evaluated. These studies mainly relied on technological developments, and physical dimensions of interior space of cars were focused on. For example, virtual vehicle interiors were rendered using CAVE display and evaluated [2]. The second research stream belonged to the studies of user's perceptions on physical attributes of interior space of cars. These studies dealt with user's perceptions and satisfaction about interior space of cars, but still relied on the physical attributes [1]. The third research stream employed emotional measures to evaluate the interior image of cars. For example, Kansei engineering was applied to evaluate vehicle interior image [3]. These studies used emotional measures for evaluation, but targeted only on several parts of vehicle interior images, instead of considering the whole psychological space of vehicle interior.

In this study, we focus on the whole interior space of the motor vehicles as a psychological space, and aim to investigate factors that represent in-vehicle roominess in psychological dimensions.

2 Methods

In order to find psychological dimensions or factors that represent the interior roominess of vehicles, we employed the two-staged approach: First, we collected psychological expressions (i.e., adjectives) that described the feelings of 'in-vehicle roominess' through brain-storming. Second, we conducted experiments to measure the in-vehicle roominess based on the psychological expressions developed from the first stage.

2.1 Psychological Expressions for In-Vehicle Roominess

Before selecting the appropriate psychological dimensions or factors that represent the interior roominess of vehicles, at the first step, we need to collect psychological expressions (i.e., adjectives) that describe the feelings of 'in-vehicle roominess' as many as possible. Seven experts (2 HCI experts, 2 car-related experts, 3 interior design experts) used the brain-storming method to collect the psychological expressions related to the feelings of 'in-vehicle roominess' without time limitation.

At the second step, the collected expressions were elaborated by the experts grouping similar expressions and eliminating irrelevant expressions. From these two steps, we could collect appropriate psychological expressions enough to find psychological dimensions or factors that represent the interior roominess of vehicles at the next stage.

2.2 Experiments for Measuring In-Vehicle Roominess

We conducted experiments to measure the in-vehicle roominess based on 105 psychological expressions developed from the first stage. In the experiments, 15 participants took part in measuring the interior roominess of vehicles. The participants were 23~29 years old (mean: 24.5, stdev.: 1.64), and 12 males and 3 females. They had 0~10 years of driving experience (mean: 4.4, stdev.: 4.07). Seven motor vehicles were used for the experiments: Benz S500L, Benz E350, BMW 5, Nissan Maxima, Nissan Altima, Hyundai Genesis and Lexus ES. These motor vehicles were selected on the basis of diversity of psychological space. Seven experts evaluated these motor vehicles, and concluded that they gave diversified impressions of interior roominess.

Every participant was asked to answer the level of agreement, based on 7-point scale, on each of 105 psychological expressions and overall in-vehicle roominess, after experiencing interior roominess of each of seven vehicles from the driver seat. The order of experiencing interior roominess of vehicles was randomized.

3 Results

We collected 105 psychological expressions (i.e., adjectives) that described the feelings of 'in-vehicle roominess' through brain-storming by 7 experts, and during the experiments for evaluating interior roominess of vehicles, based on 105 psychological expressions and overall in-vehicle roominess, each of 15 participants evaluated each of 7 motor vehicles. The data from the experiments was analyzed using factor analysis and multiple linear regression.

Factor analysis was used to find psychological dimensions or factors for representing in-vehicle roominess from the collected 105 psychological expressions, and multiple linear regression was employed to select factors that influenced in the in-vehicle roominess most.

3.1 Factor Analysis

We conducted the principal component factor analysis with varimax rotation and derived 10 factors, such as 'space completeness', 'freshness', 'openness', 'softness', 'dullness', 'narrowness', 'heaviness', 'comfort', 'closeness' and 'perspective', as shown in Table 1. 63.6% of total variance from the data was explained by 10 factors.

'Space completeness' includes such feelings of interior roominess as well-balanced, harmonious, enough, impressive, delicate, elegant, and so on. 'Freshness' represents feelings such as bright, neat, cute, magnificent, dynamic, lively, rhythmic, and so on. 'Openness' includes feelings such as high, vast, big, refreshing, pleasant, cleared, and so on. 'Softness' represents feelings such as tender, smooth, round, and so on. 'Dullness' includes feelings such as blunt, lumpy, voluminous, stumpy, and so

Table 1. Factors and their internal consistency indices

Factors	Cronbach's alpha
Space completeness	0.86
Freshness	0.80
Openness	0.84
Softness	0.74
Dullness	0.76
Narrowness	0.73
Heaviness	0.53
Comfort	0.59
Closeness	0.70
Perspective	0.70

on. ‘Narrowness’ represents feelings such as cramped, somber, stuffy, and so on. ‘Heaviness’ includes feelings such as weighty, awakening, square and level, and so on. ‘Comfort’ represents feelings such as cozy, nice, slim, friendly, conventional, and so on. ‘Closeness’ includes feelings such as dense, complex, rugged, and so on. ‘Perspective’ represents feelings such as deep, long, far-away, and so on.

Table 1 shows internal consistency of each factor’s measures based on Cronbach’s alpha index. All factors except ‘heaviness’ and ‘comfort’ indicated acceptable levels (above 0.7) of internal consistency in their measures. The measures for ‘heaviness’ and ‘comfort’ factors, however, need to be investigated further in terms of reliability.

3.2 Multiple Linear Regression

We conducted multiple linear regression with overall in-vehicle roominess as a dependent variable and the above 10 factors as independent variables. Based on the step-wise selection method, three factors were selected as the factors that influence in the in-vehicle roominess most: ‘space completeness’, ‘narrowness’ and ‘dullness’ as shown in Table 2. The linear regression line with these three factors explains about 52% of the total variability from data.

As shown in Table 2, ‘space completeness’ and ‘dullness’ have positive relationship with the overall in-vehicle roominess, whereas ‘narrowness’ have negative relationship with the overall in-vehicle roominess.

Table 2. Results of multiple linear regression

Factors	Parameter Estimates	Significance
Space completeness	0.56	p<.0001
Narrowness	-0.41	p<.0001
Dullness	0.20	p=0.0692

4 Conclusions and Discussion

We derived 10 factors for representing the in-vehicle roominess, and found 3 most influential factors in the in-vehicle roominess. ‘Space completeness’ represents how harmonious/well-balanced the elements in the interior space of vehicles are, and ‘narrowness’ represents how cramped/stuffy the elements in the interior space of vehicles are, and ‘dullness’ represents how blunt/voluminous the elements in the interior space of vehicles are. So we can expect that the in-vehicle roominess would be increased if the elements in the interior space of vehicles are located in harmonious/well-balanced ways, are not cramped/stuffy, and give blunt/voluminous feelings to drivers. Therefore, we can measure the in-vehicle roominess as a psychological space through the above factors, and we can utilize those factors to design the new interior space of the motor vehicles and/or to compare the alternative designs of the motor vehicle interior space in terms of the in-vehicle roominess.

For further study, we need to investigate how the above three most influential factors can be applied to designing sub-interior spaces of vehicles, in the practical aspects. And also, the relationships between the above three most influential factors and the physical dimensions of interior space in vehicles need to be investigated.

Acknowledgments. This study was supported by the research fund from Hyundai & Kia Corporate.

References

1. Petiot, J.-F., Salvo, C., Hossoy, I., Papalambros, P.Y., Gonzalez, R.: A Cross-Cultural Study of Users' Craftsmanship Perceptions in Vehicle Interior Design. International Journal of Product Development 7, 28–46 (2009)
2. Smith, R.C.: Shared Vision. Communications of the ACM 44, 45–48 (2001)
3. Tanoue, C., Ishizaka, K., Nagamachi, M.: Kansei Engineering: A Study on Perception of Vehicle Interior Image. International Journal of Industrial Ergonomics 19, 115–128 (1997)

Analysis of Low-Floor Bus Passengers' Behavior Patterns Using Video Observation

Ji Yeon Kim[1], Hwan Hwangbo[1], Beom Suk Jin[1], Bong-Ha Hwang[2], Young Joo Moon[2], and Yong Gu Ji[1]

[1] Department of Information & Industrial Engineering, Yonsei University, Seoul, Korea
[2] Commercial Vehicle Project Team, Hyundai Motor Company, Wanju, Korea
{jykim29,hwan976,kbf2514jin,yongguji}@yonsei.ac.kr,
{yibeizi,myjoo}@hyundai.com

Abstract. Low floor buses have regarded as a transportation which is applied to the concept of universal design. However, studies on low-floor buses in terms of universal design were rarely conducted. Moreover, passengers using low-floor buses have felt that these buses are not convenient, satisfactory and safe enough. In this study, we conducted preliminary surveys to investigate the interior design factors which affect to passengers' convenience and satisfaction. Next, we draw passengers' behavior patterns based on context of use by conducting video observation and based on these, some design suggestions were drawn to investigate the way to ensure convenience, satisfaction and safety in using low-floor buses.

Keywords: Low-floor bus, Behavior Pattern, Video Observation, Context of use, Universal design.

1 Introduction

In recent years, as the numbers of people who have physical limitations (e.g. the elderly and disabled) increased, the needs for accessibility of public facilities for those people increased as well [2].Thus, Universal design has been spreading out in public facilities and applied even to buses. Low-floor buses provide easier access to the elderly, children, and physically disabled passengers [10]. Thus, these are regarded as an effective transportation to those people and introduced in many countries (Europe, America, Japan, etc.). Korea is also gradually expanding the introduction of low-floor buses.

Despite of the trend, studies on low-floor buses that consider universal design are rarely conducted. Also, low-floor buses are not satisfying mostly to passengers including elderly people, children, and physically impaired passengers [10].

In this paper, we will try to find common behavior of passengers using low-floor buses and extract the passengers' behavior patterns. Based on this, we will suggest how to design the interior design factors of low-floor bus to improve the convenience and satisfy all passengers.

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 391–400, 2011.

2 Related Work

2.1 Universal Design in Public Transportation System

Sandhu et. al. [3] defines universal design as "Universal design has a concept that extends to a broad diversity of users who have to interact with the built environment" (p.3.3). So, many designers and architectures have to consider the needs of diverse users by using UD (Universal design) [4].

In public transportation system's design, UD should be especially considered because of the characteristics of public transportation system which is used by diverse users. Schmocker et. al. [5] indicated that elderly people living in London use buses for shopping with the highest portion compared with other public transportation system. The main reason is the increase in low-floor buses since 2005[5]. In many countries like United Kingdom, accessibility of elderly and disabled people who has lower-mobility has been increased with the introduction of low-floor buses [11]. However, Audirac [12] indicated that "low-floor buses are not UD solutions to micro-level transit accessibility although they are a form of inclusive design" (p.10). Thus, this study focuses on the low-floor buses as a research target.

2.2 Context of Use

Dey [6] defines context as "any information that can be used to characterize the situation of an entity" (p.5). Dey refers to "entity as a person, place or objects that is considered relevant to the interaction between a user and an application, including the user and applications themselves" (p.5). Maguire [7] emphasizes context of use when designers and manufacturers are developing or assessing a product in terms of human factors (HF) and extracted the component of context of use. It composed of System report and stakeholder analysis (System or product report, Stakeholder report, User/stakeholder type), user (user name, experience, knowledge and skills, personal attributes), Task (Task list), Environment (Technical environment, Physical environment, Organizational environment).

3 Methods

This study conducted a user study using video observation as well as survey. The video observation as direct observation in the field has the advantage of understanding context of user activity [10]. Thus, in order to draw passengers' behavior patterns based on context of use on using low-floor buses, video observation method was used. Before the video observation, preliminary survey was conducted to investigate the interior design factors which affect to passengers' satisfaction and convenience.

3.1 Survey

The design of the questionnaire. Before the development of questionnaire, by FGI, a bus layout was divided into boarding/getting off/front/ rear space and space for disabled passenger in wheelchair and 10 issued interior design factors were selected. Table 1 indicates the selected interior design factors based on each space.

Table 1. Interior design factors

Space	Interior design factor
A boarding space	Stanchion pipe
	Bus LED display screen
A space of getting off	Anti slipping pad
A front space	Seat on wheel pan
	Stanchion pipe
	Passenger seat
	Seat handle
	Hand rail
	Hand rail Ring
	Route map
A rear space	Seat on wheel pan
	Passenger seat
	Stanchion pipe
	Seat handle
	Hand rail
	Hand rail ring
	Route map
A space for disabled passenger in wheelchair	Foldaway seat

The questionnaire was made up of 3 parts (a total of 27 items) that included respondent demographics (3 items), level of overall satisfaction and convenience of low-floor buses (2 items) and level of satisfaction and convenience of using of each interior design factor (22 items). All items were measured using a 7-point Likert scale ranging from 'disagree strongly' to 'agree strongly'.

Subjects and Procedure. A total of 208 people including 117 females (56.2%) and 91 males (43.8%) participated. The majority were 20-29 age group which composed 41% of the samples, 10% were 10-19 age group, 15% were 30-39 age group, 13% were 40-49 age group, 14% were 50-59 age group and 7% were above 60 age group. The average height of men and women were 173.3cm and 160.5cm. Both men and women showed an error of approximately 3cm considering the average height of Korean.

The preliminary survey was conducted for 6 days. It targeted the passengers who were waiting for the low-floor buses at a bus station to reflect earlier experiences and who were on the low-floor buses. The questionnaire data were analyzed with SPSS statistics version 18.0 and ANOVA (α=0.05) was used.

3.2 Video Observation

Data collection. In this study, data were collected by two low-floor buses which were manufactured by Hyundai motor company. We collected data during 14days including weekend, weekdays, and rainy days.

Helen et.al [9] indicate that researchers who use data gathering method with video have to be careful of Hawthorne effect and shouldn't miss the ongoing things out of the

Fig. 1. The examples of collected data

Table 2. Modified components of context of use

Category	Sub-Category	Component
User	User type	Passengers who do not have physical limitations
		Passengers who have physical limitations
	User characteristic	The level of personal belongings
		Sex
Task	Task list	Boarding
		Card tagging
		Movement in the inside
		Supporting passenger's body
		Sitting
		Standing
		Caring personal belongings
		Controlling the side widow or air conditioner
		Identifying the route map
		Pressing the call buzzer
		Getting off
		Unexpected incident
Environment	Technical	Card reader
		Door handle
		Seat handle
		Foldaway seat
		Wheelchair fixing device
		Seat on wheel pan
		Steps in rear aisle
		Hand rail
		Hand rail ring
		Stanchion pipe
		Call buzzer
		Side window
		Route map
		Air conditioner
	Physical	Rainy weather
	Organizational	The level of congestion

camera view. Thus, this study used a total of 8 cameras (4 cameras per bus) at various angle and which were set up at ceiling to avoid becoming aware of passengers.

Data Analysis. The collected data were analyzed by conducting 2 steps by three researchers. First step, the collected data were recorded on an event basis.

Fig. 1. were captured by viewer program and showed passengers' behavior in two low-floor buses. Four camera's data were showed all together and if necessary, one camera's data can be shown separately.

Second step, recorded events were shown repeatedly and finally analyzed based on modified components of context of use. By conducting FGI, we extracted the component of context of use in low-floor buses by borrowing and revising Maguire's concept [7] depending on characteristic of low-floor buses (Table 2). To gather well-reflected components under context of low-floor buses, a total of 5 participants (3 of HCI researchers, 2 of low-floor bus developers attended.

The object of this study is not concentreated on comparison of product of stakeholders. Thus, we emitted category of system report and stakeholders. Moreover, video observation focuses on the patterns of people by observing how people interact with one another and their physical environment [9]. Thus, User categories were divided into User type (Passengers who do not have physical limitations, Passengers who have physical limitations) and User characteristics (The level of personal belongings and sex). Regarding passengers' behavior in the low-floor buses, 12 tasks and 14 interior design factors were selected as a task and as technical environment. Rainy weather was considered as a physical environment because temperature change could not be reflected. The reason was that this study was conducted in summer. Finally, passengers' behavior can be changed according to the level of congestion [8]. Thus, the level of congestion was considered as an organizational environment.

4. Results

4.1 Survey Results

Comparison between groups, by sex, height and age, was conducted by ANOVA. ANOVA analysis showed that a total of 27 items were analyzed and 5 items were significant. There were significant differences among groups by sex and age. There were no significant differences of height. Table 3 indicates the specific relationship between significant items and passenger's sex, height and age.

The results demonstrate that passengers' sex and age are important factors to improve overall convenience of low-floor buses. Moreover, significant differences between male and female can be related to frequency of use of these interior design factors in the convenience of handrail and handrail ring. These assumptions were verified by observing passengers' behavior through video observation method.

Table 3. Summary of statistically significant results

Item	Passengers' characteristic		
	Sex	Height	Age
Overall convenience of using low-floor buses	F=6.351*	-	F=4.074**
Overall satisfaction of using low-floor buses	-	-	F=3.581**
Convenience of holding handrail	F=4.042*	-	-
Satisfaction of using handrail	F=6.313*	-	-
Convenience of holding handrail ring	F=4.793*	-	-

*Note: *: $p < 0.05$, **: $p < 0.01$, -: non-significant.

4.2 Video Observation Results

Events analysis. An analyzed total of 996 events contained 500 female events, 472 male events, and 24 were related to group events (Table 4). Group events means that the passengers in low-floor buses get all tangled up together because many passengers are in buses at the same time, so, more than 2 passengers were considered as same event. As passengers who have physical limitations, the elderly were mainly considered to draw passengers' behavior patterns. Because children and physically disabled passengers were rarely observed and pregnant women were not at all observed.

Table 4. The type of passengers

Passengers' type	Male	Female	Group	Total
Passengers who don't have physical limitations	373	369	22	764
The elderly	78	122	0	200
Physically disabled passengers	4	0	0	4
Children	17	9	2	28
Pregnant women	0	0	0	0
Total	472	500	24	996

Moreover, events were analyzed by the level of personal belongings (Table 5). The level of personal belongings was divided into 3 levels which were nothingness, light and heavy. Light personal belongings such as purse and hand bag indicated that it doesn't affect passengers' movements. Considering that the passenger's personal belongings affect his or her behavior, we classified as heavy personal belonging. Female passengers carried personal belongings with high frequency compared to male passengers and most of the passengers carried light personal belongings.

Passengers' Behavior Patterns. A total of 16 behavior patterns were drawn and Table 6 gives a summarized overview.

Table 5. The level of personal belongings

The level of personal belonging	Male	Female	Total
Nothingness	215	76	291
Light	236	375	611
Heavy	21	49	70
Total	472	500	972

Figure-2 indicates the behavior pattern of male passengers who are on board standing and do not have physical limitations.

Based on analyzed passengers' behavior patterns, we categorized passengers' behavior patterns according to passengers' sex, type and context.

First, in the case of male passengers who do not have physical limitations, they used support-purposed interior design factors such as seat handle when they sit and stand. However, in case there were not exist support-purposed interior design factors in their surrounding, they tended to grip head of seats, window frames and wheelchair fixing device et al. When male passengers who do not have physical limitations were standing, they tended to use handrail more frequently than female passengers. Moreover, they were able to use this interior design factor easily in the context of congestion.

Next, in the case of female passengers who do not have physical limitations, their needs for support-purposed interior design factors were similar to male passengers' needs. However, female passengers tended to use relatively low located interior design factors when they were standing. Thus, they showed unsteady behavior in the context of congestion because interior design factors to support female passengers were relatively not enough.

Table 6. Summary of passengers' behavior patterns

No	Type of behavior pattern
1	Sitting, Male passengers who do not have physical limitations
2	Standing, Male passengers who do not have physical limitations
3	Rainy weather, Male passengers who do not have physical limitations
4	Congestion, Male passengers who do not have physical limitations
5	Sitting, Female passengers who do not have physical limitations
6	Standing, Female passengers who do not have physical limitations
7	Female passengers who carry heavy personal belongings and do not have physical limitations
8	Congestion, Female passengers who do not have physical limitations
9	Sitting, Male passengers who have physical limitations
10	Standing, Male passengers who have physical limitations
11	Male passengers who carry heavy personal belongings and have physical limitations
12	Congestion, Male passengers who have physical limitations
13	Sitting, Female passengers who have physical limitation
14	Standing, Female passengers who have physical limitations
15	Female passengers who carry heavy personal belongings and have physical limitations
16	Congestion, Female passengers who have physical limitations

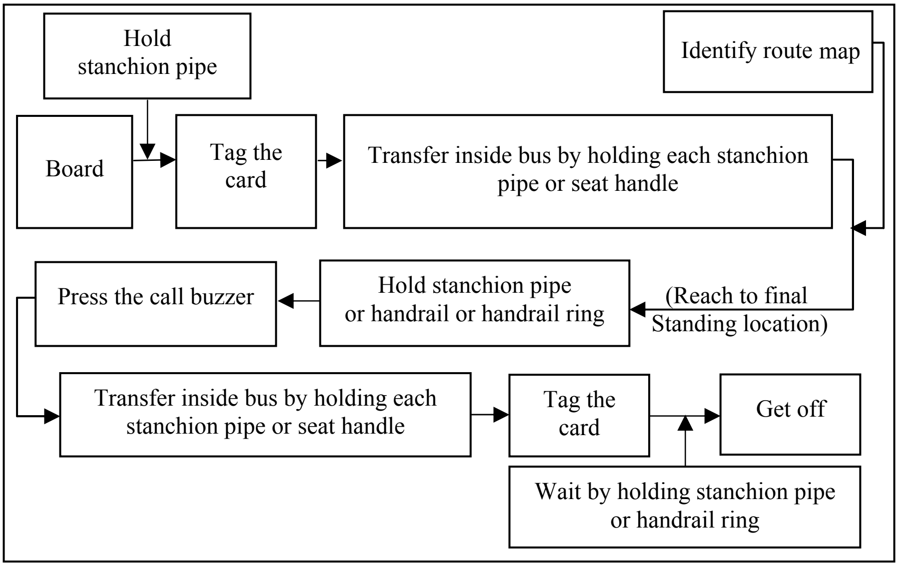


Fig. 2. One example of passengers' behavior patterns

Third, male passengers who have physical limitations showed similar behaviors using support-purposed interior design factors compared to those of male passengers who do not have physical limitations. However, they relied on support-purposed interior design factors from boarding to getting off and they tended to use seats in the front space rather than seats in the rear space.

Finally, In the case of female passengers who have physical limitations indicated high frequency for holding interior design factors and caring personal belongings. Moreover, they tended to use seats in the front space like male passengers who have physical limitations and support themselves with the door handle when they were boarding the buses. When female passengers who have physical limitations were standing, they showed similar behavior patterns compared to those of female passengers who do not have physical limitations. However, their mobility were reduced when they were moving to the rear space because of the steps in the rear aisle

In general, passengers showed low frequency of using rear aisle in the bus when it was crowded and there was a tendency to wait to get off in the getting off space. Furthermore, when sitting in the space for disabled passenger in wheelchair, passengers often held wheelchair fixing device on the wall. In the case of use of handrail ring, passengers supported themselves with an additional interior design factor like stanchion pipe and handrail because of swing.

5 Design Suggestion

The results of passengers' behavior patterns gave us insights on how to design low-floor buses. The results indicated that some interior design factors had to be improved in universal designs. Because of the steps in the rear aisle, the passengers who have physical limitations used the rear space with less frequency and the front space was relatively crowded with passengers in the context of congestion. Thus, designers and manufacturers of low-floor buses can consider that they should get rid of steps in the rear aisle. However, they have to keep in mind that non-step in the rear aisle could lead to more difficulty in sitting on the seats on rear wheel pans. Moreover, the front space has to be designed to accommodate passengers who have physical limitations.

In the case of seats on wheel pan which were located in the front and rear spaces, many passengers had difficulty sitting because of the high height. In this case, additional support-purposed interior design factors can be added around the seats.

Finally, the heights of handrail and handrail rings also have to be considered when low-floor buses are designed.

6 Conclusions

In this study we point out the most interesting findings regarding behavior patterns of passengers based on context of use of low-floor buses. As a consequence of this study, there are some limitations. First, passengers' behavior patterns which were demonstrated in this study can be changed according to overall layout and characteristic of interior design factors in low-floor buses. Second, as passengers who have physical limitations, pregnant women could not be considered. This is because pregnant women were not observed in buses which were set up the cameras for the purpose of study in our study period. Given these limitations, user testing can be conducted with diverse users including pregnant women.

However, this study shows significance in the consideration of using video observation in dynamic circumstance and importance of context of use. Additionally, it can be applied to the passengers' behavior patterns based on various vehicles as a methodology, and we can anticipate that this study will be utilized as the guideline for next generation low-floor bus design.

References

1. Levine, J.C., Torng, G.W.: Dwell-Time Effects of Low-Floor Bus Design. J. Transp. Engrg. 120, 914–929 (1994)
2. Froyen, H., Verdonck, E., Meester, D.D., Heylighten, A.: Documenting handicap situations and eliminations through universal design patterns. Australasian Medical Journal 1, 199–203 (2009)
3. Sandhu, J.S.: Universal Design Handbook: Chapter 3 An Integrated Approach to Universal Design: Toward the Inclusion of All Ages, Cultures, and Diversity. McGraw-Hill, New York (2001)
4. Scott, S.S., Mcguire, M.J., Shaw, S.F.: Universal Design for Instruction: A New Paradigm for Adult Instruction in Postsecondary Education. Remedial and Special Education 24, 369–379 (2003)
5. Schmocker, J.D., Quddus, M.A., Noland, R.B., Bell, M.G.H.: Mode choice of older and disabled people: a case study of shopping trips in London. Journal of Transport Geography 16, 257–267 (2008)

6. Dey, A.K.: Understanding and Using Context. Personal and Ubiquitous Computing 5, 4–7 (2001)
7. Maguire, M.: Context of Use within usability activities. Int. J. Human-Computer Studies 55, 453–483 (2001)
8. Hennessy, D.A., Wiesenthal, D.L.: Traffic congestion, Driver stress and Driver aggression. Aggressive Behavior 25, 409–423 (1999)
9. Helen, S., Yvonne, R., Jenny, P.: Beyond human computer interaction, 2nd edn. John Wiley & Sons Ltd., West Sussex (2006)
10. Transportation Research Board National Research Council.: TCRP Report 41: New Designs and Operating Experiences with Low-Floor Buses. National Academy Press, Washington, D.C (1998)
11. Transportation Research Board, `http://www.trb.org`
12. Audirac, I.: Accessing Transit as Universal Design. Journal of Planning Literature 23, 4–16 (2008)

The Effective IVIS Menu and Control Type of an Instrumental Gauge Cluster and Steering Wheel Remote Control with a Menu Traversal

Seong M. Kim[1], Jaekyu Park[2], Jaeho Choe[3], and Eui S. Jung[2]

[1] Graduated School of Information Management and Security, Korea University, 5-1 Anam-dong, Seongbuk-gu, Seoul, Korea
iampreppie@korea.ac.kr
[2] Department of Industrial and Management Engineering, Korea University, 5-1 Anam-dong, Seongbuk-gu, Seoul, Korea
ergopark@gmail.com, ejung@korea.ac.kr
[3] Department of Industrial and Management Engineering, Daejin University, 11-1 Seondan-dong, Pocheon, Korea
jhchoe@daejin.ac.kr

Abstract. The present study investigated the effective IVIS menu and control type of the instrument gauge cluster and steering wheel remote control. Participants performed menu traversal tasks with a steering wheel remote control and gauge cluster display in a driving video simulation. Two steps of experiment were conducted. The first was focused on the menu type and within-subject factorial design was implemented with two levels of menu types (the spread and overlapped menu type), two levels of control types (the wheel and touch wheel controller) and two levels of menu traversal tasks. Subjective ratings of a preference (using modified Likert-type rating scale) and a menu traversal time and routing error were obtained as performance measures. ANOVA results showed that the menu type, control type and the interaction of the control and task were significantly affected by each of the independent variables. The result implied that the spread menu type and wheel controller were more effective. The second experiment was focused on the control type with the spread menu type and within-subject factorial design was implemented with two levels of control types (the wheel and 4-way directional controller) and two levels of menu traversal tasks. ANOVA results of a performance showed that the control type and task were significantly affected. The result implied that the wheel control type with the spread menu type was more effective IVIS interface alternative on a gauge cluster and steering wheel remote control.

Keywords: IVIS Interface, IVIS Menu Traversal, IVIS Menu Type and Control Type.

1 Introduction

In-Vehicle Information System (IVIS) is increased in the automobile market. This is the one of the main system for the vehicle. This trend is along with the electronic technology development and reflects the driver requirement for information interaction.

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 401–410, 2011.

This system (IVIS) makes the increase of the driver information interaction in the vehicle. This means that the driver workload is increased. So, to reduce the driver workload the various approaches to the IVIS interface have accomplished. These have focused on the IVIS menu and control type, location and manipulation efficiency at the center-fascia. These are the effort to provide more safe and effective interface to the driver. [1, 2, 3, 4, 5].

However, IVIS have been changed. From the center-fascia to the gauge cluster, IVIS expands to the whole driver cockpit. This is not only location changes but also information contents transfer. The information that is represented at the center fascia and gauge cluster is not clearly distinguished. However, there is common trend of information categorization in the vehicle market. While the navigation, ventilation and entertainment information are represented IVIS at the center-facia, the convenience system setting, driving information and auxiliary entertainment information are applied IVIS at the gauge cluster. The expansion of IVIS area is required the complex manipulation to the driver. More information is displayed at the gauge cluster that is made by LCD panel as like computer monitor. To control this information the driver may manipulate the steering wheel remote control that is consists of various functions. The previous steering wheel remote control is applied one function to one control. However, more information at gauge cluster is required more complex steering wheel remote control as like 4-way directional, wheel, and touch wheel controller.

These complex manipulations with information display at the gauge cluster and control at the steering wheel makes more workload to the driver also. Especially, as the location of IVIS complex remote control is changed from center console to steering wheel, not only steering control but also menu manipulation are required to the driver.

When there was the change of the display area and input device constraint, other studies that were conducted without the same constraint were not able to apply. [6] Therefore, in this study we focused on the constraint of the gauge cluster display and steering wheel remote control. And we proposed the effective IVIS menu and control alternative through the menu traversal.

2 Methods

Other study on the IVIS interface with the gauge cluster display and steering wheel remote control was focused on the effective menu depth and breath. This study was proposed that the optimum menu structure was 3-depth and 7-breath. [7] And he proposed that the driver preferred entertainment information manipulation contents. This was the auxiliary entertainment information of the other IVIS that was located in center-fascia. And this information was consisted of the music channel selection as like radio, CD, MP3 and outer entertainment device that was kind of i-pod and portable storage device. As the aspect of the system proximity and manipulation convenience, the IVIS that was located in gauge cluster and steering wheel had more advantage.

In this study, we reflected these menu structures and information contents. And to propose the effective menu and control type, we approached to the IVIS menu traversal with the task operation. The task was consisted of two operations. The first was the steering control with the driving video and the second was the menu manipulation. These tasks were applied simultaneously. To conduct this menu traversal experiment, we set up the driving simulator. The driving simulator was consisted of the steering wheel, LCD gauge cluster, driver seat, pedal and driving video projector. The driver who sat down the driver seat saw the front gauge cluster and the driving video and manipulated the steering wheel and remote control. (Fig. 1).

The driver who was the experiment participant controlled the steering wheel with the following direction that was represented by a driving video. And the driver manipulated the remote control mock-up with the gauge cluster menu display. In this experimental environment the driver conducted the menu traversal task and we measured the preference, performance time and error for menu and control type.

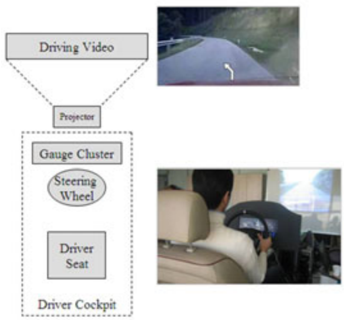


Fig. 1. The experiment environment of the IVIS menu traversal

2.1 Subjects

Two kinds of experiment were conducted. In the first experiment, nineteen Korean male and female drivers (age 40.2 years, driving experiment 13.8 years) participated in this study. In the second experiment, thirteen healthy male and female drivers (age 38.5 years, driving experiment 12.6 years) participated. All the subjects read and signed a consent form before the experiment. They were paid for their participation.

2.2 Apparatus

The gauge cluster that was displayed the IVIS menu was LCD device. And the gauge cluster had two modes. The one was driving mode. And this mode was consisted of the analog representation of the speedometer (left side) and tachometer (right side). These were similar to a commercial vehicle. The other was menu mode. And this mode activated by the wake-up control that was located in the steering wheel remote control. If the menu mode was 'On', the right side representation of the analog type tachometer was changed to the IVIS menu. (In this mode, the tachometer representation type was changed from the analog to the digital.) (Fig. 2).

The menu alternative was encoded by the Flash Program and the main computer displayed this alternative to gauge cluster. The steering remote control was connected by USB channel to the main computer. So, the gauge cluster was the output display

Fig. 2. The driving mode (left) and menu mode (right) of the gauge cluster

device and the steering wheel remote control was the input control device. The performance time and error of the menu traversal was encoded by the Flash Program, also. As the results, this time and traversal error was recorded by the software. The driving video projector was connected by the sub-computer.

2.3 Experimental Design

Two kinds of experiment were conducted. The first experiment was focused on the effective menu type with rotation movement control. The wheel and touch controller that is not applied to a vehicle but applied to an electronic device were used for the control type alternative. These controllers were known to the effective device for manipulating amount of information.

The second experiment was the comparison the rotation movement control to the 4-way directional movement control on the basis of the optimal menu type that was selected by the first experiment. The 4-way directional controller was applied to a vehicle and used to other IVIS study. [1, 2] So, the second experiment was focused on the comparison a new type controller with an existing controller.

2.3.1 Experiment for the Effective Menu Type with Two Controls

To measure a preference, performance time and routing error as dependent variables, we used a within-subject factorial design with two levels of control types, two levels of menu types and two levels of menu traversal tasks as independent variables.

Two levels of control types were a wheel controller and touch wheel controller. Two types of controller had a same movement direction. However, they were different to a movement mechanism. A menu movement was conducted by a clock-wise or a counter clock-wise manipulation. And a menu selection was performed with 'OK' control that was located in the center of a controller. (Fig. 3).

Two levels of menu types were a spread menu and overlapped menu. If a lower depth menu was selected, the spread menu was displayed a lower depth menu with an upper depth menu. On the other hand, the overlapped menu was displayed a just lower depth menu. (An upper depth menu was disappeared.) These menus were used to many electronic devices. A round or circle menu shape was applied for the rotation movement control, also. (Fig. 4).

Fig. 3. A wheel controller (left) and touch wheel controller (right)

Fig. 4. A spread menu type (left) and overlapped menu type (right)

Two levels of menu traversal tasks were a task 1 and task 2. A task 1 was consisted of a driving and convenience device setting. And a task 2 was a menu traversal for music channel setting. (Table 1) So, we made eight alternatives with control types, menu types and tasks. (Table 2)

Table 1. Two types of menu traversal tasks

Task 1	Wake up → Driving Environments → LDWS → OFF → return to first menu
Task 2	Wake up → AV → CD → Track no. 10 → return to first menu → Media → M-Station → Track no. 23

Table 2. An alternative for an IVIS menu traversal experiment

Control Type	Menu Type	Task	Alternative
Wheel	Spread	1	1
		2	2
	Overlapped	1	3
		2	4
Touch Wheel	Spread	1	5
		2	6
	Overlapped	1	7
		2	8

2.3.2 Experiment for the Effective Control Type

To measure a preference, performance time and routing error as dependent variables, we used a within-subject factorial design with two levels of control types and two levels of menu traversal tasks as independent variables. Two levels of control types were a wheel and directional controller. (A wheel controller was more effective in the first experiment.) (Fig. 5).

And a menu type that was selected by the first experiment was used. The task was same to the first experiment, also. So, we made four alternatives with control types and tasks (Table 3).

Fig. 5. A wheel controller (left) and 4-way directional controller (right)

Table 3. An alternative for an IVIS menu traversal

Control Type	Task	Alternative
Wheel	1	1
	2	2
Directional	1	3
	2	4

2.4 Experimental Procedure

A participant conducted a pre-test for a menu traversal and read the menu traversal task flow and memorized. Then, he or she participated in the experiment. Under the steering wheel control task with following direction that was represented by a driving video, the participant performed the menu traversal task. The steering wheel control error was recorded by an operator. If the participant failed to steering control more than two times, the experiment was conducted again. Whenever the participant ends a menu traversal for each alternative, a preference was measured by an operator with Likert-type 100-scale and a performance time and routing error was recorded by the software.

3 Results

3.1 Preference and Performance Analysis of the Effective Menu Type

The ANOVA results for the menu traversal preference showed significant main effects of control type, $F(1, 18) = 12.210$, $p = 0.003$ and menu type $F(1, 18) = 6.617$. The interaction of the control type and task, $F(1, 18) = 4.983$, $p = 0.039$ was significant, also. (Table 4). The result showed that the wheel controller and the spread menu type were more preferred. (Fig. 6).

Table 4. ANOVA results for the preference

Source	DF	SS	MS	F-Value	Pr > F
Control type	1	3971.901	3971.901	12.210	0.003*
Menu type	1	469.007	469.007	6.617	0.019*
Task	1	49.796	49.796	0.179	0.678
Control type × Menu type	1	84.007	84.007	0.590	0.452
Control type × Task	1	703.480	703.480	4.983	0.039*
Menu type × Task	1	123.480	123.480	0.852	0.368
Residual	18	1570.211	87.234		

**: significant at α = 0.05.*

When there is no interaction with independent variables, the result of the mean difference test between the main effect and each level of the independent variable is reliable. However, in this experiment, the interaction of control type and task were significant. This showed that the task was affected by the control type and this interaction was able to have effect on the result of mean difference between the main effect and each level of the independent variable.

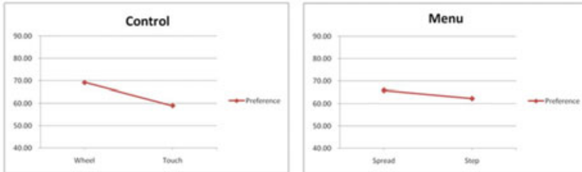


Fig. 6. The means of the control (left) and menu type (right) *(Unit: rating score)*

This interaction showed at Fig. 7. The wheel control kept better preference regardless of the task. However, the preference of the touch wheel control was decreased at the music channel manipulation task.

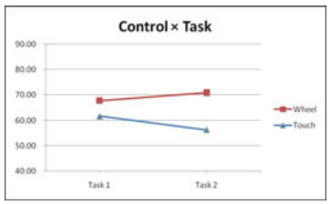


Fig. 7. The means of the interaction *(Unit: rating score)*

Table 5. ANOVA results for the performance time

Source	DF	SS	MS	F-Value	Pr > F
Control type	1	2846.716	2846.716	21.887	0.000*
Menu type	1	10.109	10.109	0.254	0.621
Task	1	35178.906	35178.906	186.277	0.000*
Control type × Menu type	1	122.401	122.401	1.066	0.316
Control type × Task	1	3052.852	3052.852	26.191	0.000*
Menu type × Task	1	91.916	91.916	1.143	0.299
Residual	18	756.601	42.033		

**: significant at* $\alpha = 0.05$.

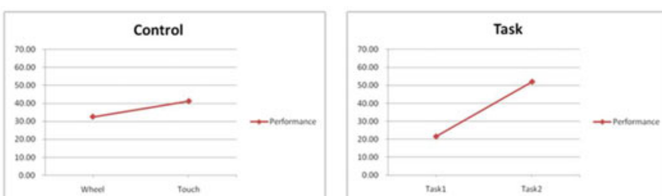


Fig. 8. The means of the control type (left) and task (right) *(Unit: sec.)*

Table 6. The routing error frequency of the controller

Control Type	Wheel Controller	Touch Wheel Controller
Routing Error Frequency	0.65	1.63

The ANOVA results for the menu traversal performance showed significant main effects of control type, F(1, 18) = 21.887, p = 0.000 and task, F(1, 18) = 186.277, p = 0.000. The interaction of the control type and task, F(1, 18) = 26.191, p = 0.000 was significant, as like the preference analysis. (Table 5). The result showed that the performance time and the routing error of the wheel controller was better. (Fig. 8 and Table 6) And the performance time of the task 1 was better, also. (Fig. 8).

As similar to the preference analysis result, there was the interaction of the control type and task. (Fig. 9) In the task 2, the difference between the wheel and the touch wheel controller performance was more. As a result of the preference and performance analysis, the effective IVIS interface alternative was the spread menu and wheel controller.

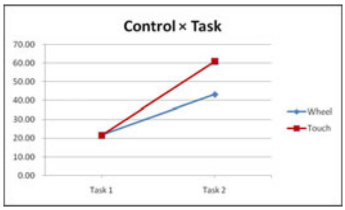


Fig. 9. The means of the interaction with the control type and task *(Unit: sec.)*

3.2 Preference and Performance Analysis of the Effective Control Type

In the second experiment, we compared the wheel controller with the 4-way directional controller. The ANOVA results for the menu traversal preference showed that there was no significant effect. (Table 7)

The ANOVA results for the menu traversal performance showed significant main effects of task, F(1,12) = 71.967, p = 0.000 at $\alpha = 0.05$ and control type, F(1, 12) = 3.929 at $\alpha = 0.1$. The interaction of the control type and task was not significant. (Table 8).

Table 7. ANOVA results for the preference

Source	DF	SS	MS	F-Value	Pr > F
Control	1	12.019	12.019	0.028	0.869
Task	1	0.481	0.481	0.003	0.954
Control × Task	1	300.481	300.481	1.453	0.251
Residual	12	2480.769	206.731		

**: significant at α = 0.05.*

Table 8. ANOVA results for the performance time

Source	DF	SS	MS	F-Value	Pr > F
Control	1	122.769	122.769	3.929	0.071**
Task	1	6705.582	6705.582	71.967	0.000*
Control × Task	1	27.623	27.623	0.994	0.338
Residual	12	333.454	27.788		

**: significant at α = 0.05.*
***: significant at α = 0.1.*

The result showed that the performance time of task 1 and the wheel control was better. However, the routing error frequency was similar. (Fig. 10 and Table 9).

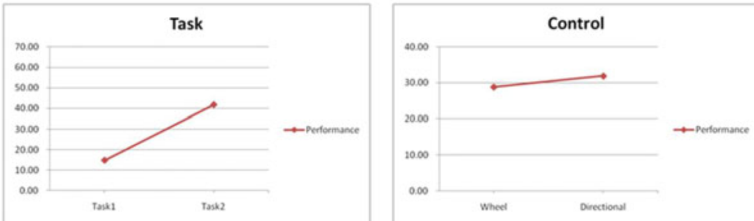


Fig. 10. The means of the task (left) and control type (right) *(Unit: sec.)*

Table 9. The routing error frequency of the controller

Control Type	Wheel Controller	Directional Controller
Routing Error Frequency	0.60	0.65

4 Discussion

This study was intended to propose the effective IVIS interface of the gauge cluster and steering wheel remote control with menu traversal. Present study considered two steps of experiments. In the first experiment, we proposed the effective IVIS menu type with two control types and tasks. Two types of controller had the same movement direction. However, they were different to the movement mechanism. The results of ANOVA showed that the spread menu type was better preference than the overlapped menu type that was a circle type of shape in similar to the controller movement direction. However, the difference of the performance time was not significant. The spread menu type had an advantage on the menu traversal. Because each menu traversal step was represented and confirmed, it was supposed that this menu type was more preferred than the overlapped menu type. To consider the menu traversal in the driving situation, it was supposed that the effective menu type should be provided the confirmation of the menu traversal step to the driver. In two types of controllers, the result of ANOVA showed that the wheel control type was more preferred and had better performance times. To consider the blind control in the driving situation, it was supposed that the wheel control that had the mechanical feedback more advantage than the touch wheel control type that had not the manipulation feedback. This difference was significant in the condition of the task. There was no difference at the task 1 that was simple menu manipulation. However, at the task 2 that was more complex menu manipulation, the wheel controller had more advantage on the preference and performance time. Therefore, it was supposed that the wheel control was more effective controller when the driver manipulated the amount of information.

In the second experiment, we proposed the effective IVIS control type. The result of ANOVA showed that the wheel control type had better performance time than the 4-way directional control type. However, the difference of the preference was not significant. It meant that the wheel controller had better performance and similar preference with the 4-way directional controller that was used to some vehicle. Therefore, it is supposed that the wheel controller was more effective control alternative.

This study proposed that the effective IVIS interface was the spread menu type and wheel control type. However, there were some constraints in this interface. First, present study was focused on the right-hand control. Though a study of a effective location of IVIS menu control that is on a left side or right side of a steering wheel is required, to consider the steering wheel design feature of symmetry, a study of a interaction between left-hand and right-hand control will be required, also. Another constraint was the age. The age of the participant was thirties to fifties. These had less experience to manipulate a touch wheel control that was applied to other electronic devices than twenties. Therefore, a study that was focused on the twenties who is familiar with a touch wheel control will be required.

References

1. Lim, H.U., Jung Eui, S., Park, S., Jeong, S.W.: Comparisons of Driver Performance with Control Types of the Driver Information System. Journal of The Ergonomic Society of Korea 26(1), 1–10 (2007)
2. Lim, H.U., Oh, Y., Park, S., Jung, E.S.: Research on function application in steering wheel. In: Proceedings of Korean Ergonomics Conference-Spring (2006)
3. Jung, E.S.: Development guideline of Next Generation IVIS interface. Hyundai Motor Company Technical Report (2005)
4. Galer, F.: The safe design of in-vehicle information and support systems, the human factors issues. International Journal of Vehicle Design 16, 158–167 (1995)
5. Stevens, A., Quimby, A., Board, A., Kersloot, T., Burns, P.: Design Guidelines for Safety of In-vehicle Information System. Transport Research Laboratory PA3721/01 (2002)
6. Jo, K., Choi, H., Han, K.: Search for the Efficient Hierarchical Data Structure in Mobile Screen. Journal of The Korean Society for Cognitive Science 18(2), 193–221 (2007)
7. Hong, S.P., Kim, S.M., Seong, M., Park, S., Jung, E.S.: Design of a Menu Structure for the Instrument Cluster IVIS using Taguchi Method. Journal of The Ergonomic Society of Korea 29(1), 39–46 (2010)

Assessing the Effect of a Power-Flow Gauge on Driving Behaviors Affecting Energy Consumption

Sang-Hwan Kim[1], Heramb Dandekar[2], Edgar Camez[3], and Heather Harrelson[1]

[1] University of Michigan-Dearborn
[2] Chrysler Group LLC.
[3] Ford Motor Company
dysart@umich.edu, hd150@chrysler.com, ecamez@ford.com,
hharrels@umd.umich.edu

Abstract. The objective of this study was to investigate the effects of a power-flow gauge for indicating current level of "economic" driving for a simulated electric vehicle based on drivers' acceleration and braking pedal pressing behaviors. Sixteen participants were asked to drive a driving simulator with/without integrating the power-flow gauge interface for city and highway. The mean and kurtosis (stability) of acceleration and brake pedal press positions were recorded. Results showed stable (non-aggressive) acceleration behaviors when drivers used the interface. This indicates that the presence of a power-flow gauge encouraged the drivers to maintain a relatively steady acceleration pedal position as well as how the changes in driving behavior may affect energy consumptions in forms of economic driving.

Keywords: Power-flow gauge interface, Driving behaviors, Economic-driving.

1 Introduction

With awareness rising about depleting fossil fuels, concentration on alternative energy sources for vehicles is at an all time high. Hybrid vehicles are still of some interest, but the main focus now is on electric vehicles. Due to a lack of sufficient charging stations around the world and a quickly rechargeable battery, drivers are experiencing extreme amounts of anxiety with regards to reaching their destination. As a result, it is important that drivers utilize any means of conserving energy.

For conventional internal combustion powered (gasoline engine) vehicles, historical studies have demonstrated that the modification of driving behaviors can help reduce levels of fuel consumption and emissions. This behavior is referred to as "eco-driving," a non-aggressive and smooth driving behavior. Eco-driving is thought to improve fuel economy because drivers observe the speed limit and avoid rapid acceleration [1]. Evans demonstrated that a driver could reduce fuel consumption by as much as 14% without increasing trip time by reducing acceleration levels and generally driving more 'gently' combined with a skillful avoidance of stops [2]. Waters and Laker also found fuel savings of approximately 15% when drivers were asked to drive economically [3]. In order to provide drivers with an external guide to prevent

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 411–417, 2011.

non-economic driving behaviors, Larsson and Ericsson suggested an acceleration advisory tool to present advice to drivers through resistance in the accelerator pedal when they try to accelerate rapidly [4]. The empirical results demonstrated the presence of the advisory tool significantly reduced strong acceleration behaviors and affected emissions.

There have been other approaches to affect driving behaviors leading to fuel consumption; one example is the addition of a driving interface that would give feedback regarding driving behaviors. Voort et al. demonstrated a prototype of a fuel-efficiency support tool including a human-machine interface [5]. The results of their experiment showed drivers were able to reduce overall fuel consumption by 16% compared with 'normal driving' while the same drivers were only able to achieve a reduction of 9% when asked to drive fuel efficiently without the support. A recent study conducted by the National Highway Traffic Safety Administration confirmed that the addition of an interface reduced fuel consumption in gas vehicles [6] [7]. They developed several interface prototypes (FEDI: fuel economy driver interface) to provide instantaneous feedback for drivers to be aware of if their driving behaviors are energy efficient or not. Through usability testing using static conceptual prototypes, they evaluated and investigated user interpretation of several combinations of interface features (e.g., bar, symbols, and/or text) being used today [6]. Based on the findings, horizontal bars and figure representations were most usable and text increased comprehension. They then conducted an experiment using working prototypes integrated with a driving simulator [7]. The results showed that once drivers become familiar with how the FEDI works, they would be able to adjust their driving behaviors in a manner that reduces energy consumption. In addition to this study, an automotive manufacturer recently announced that an interface using a metaphor of a glass of water indicating aggressive driving behaviors is expected to decrease fuel consumption [8].

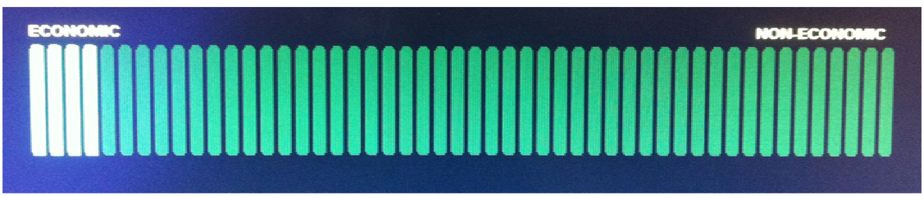
However, even though the previous studies showed effects of the interface on fuel consumption, it has been focused on conventional gas vehicles. Since not only acceleration may affect energy consumptions in electric vehicles, but braking does also (i.e., regenerative braking allows the battery to recover charge). Braking behaviors must also be examined when using such an interface. In addition, it also is necessary to investigate how the interface affects driving behaviors, such as acceleration and brake pedal pressing, which may affects energy consumption.

This study was to investigate the effect of a power-flow gauge interface on driving behaviors including acceleration and braking (potentially reducing energy consumption), in terms of human-machine interface (HMI). An experiment was conducted using a driving simulator integrated with a prototype interface. Unlike previous studies, which have focused primarily on gas vehicles, this study simulated the usage of the interface in a battery electric vehicle (BEV).

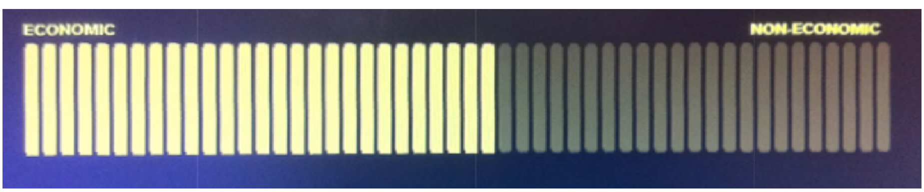
2 Methods

2.1 Prototype Interface

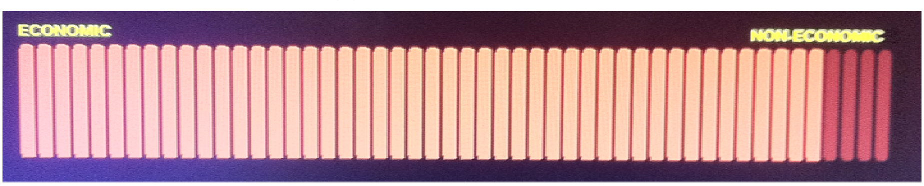
A Java-based prototype interface for displaying the power-flow gauge was developed and integrated with a driving simulator. The prototype application was designed to display the current status of economic driving to the user. The design incorporated changes of acceleration and brake pedal levels (accelerations and decelerations) and current driving speed. Aggressive and abrupt pedal presses increased the non-economic driving level and it was displayed in the interface in real-time. The graphical feature of the interface was a horizontal bar indicating the driving behavior of the driver, which was shown by Jeness et al. to be more effective [6]. A longer bar represented more non-economic driving. Along with the changes in length of the bar, the color of the indicator also gradually changed from green to red as the driver drove in a non-economic fashion (see Fig. 1 (a) through (c)). Since the interface was developed for electric vehicles, the gauge was designed to change its color to green and to show a blinking text of "Charging" when the brake pedal was pressed for stopping or slowing down the simulation vehicle (see Fig.1 (d)). In addition to this, the Java application was designed to record the values of gas and brake pedal positions as a text file. The pedal values ranged from 0 to 10000 and were recorded on every 100 msec (10 Hz).



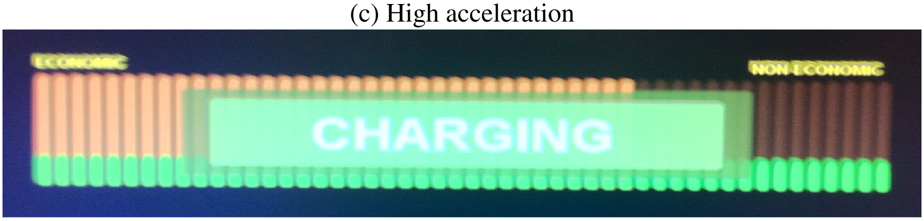
(a) Low acceleration



(b) Medium acceleration



(c) High acceleration



(d) Regenerative braking

Fig. 1. Power-Flow Gauge Conditions

Fig. 2. Driving Simulator

2.2 Experiment Setup

A driving simulator was used for the experiment, which allowed the subject to adjust the seat in two directions and the steering wheel in two directions as well. The setup included a steering wheel, a gear lever and the pedals, which are integrated with a commercial driving simulator, which simulates the driving experience. The equipment also included a projector to show driving scene and a LCD monitor to present the power-flow gauge. The Java application for the power-flow gauge was run on a PC and the gauge was displayed on the monitor in front of the driver in the place of a conventional instrument cluster. The road scene generated by the simulator software was projected on to the wall by the projector. Figure 2 shows the simulator equipment.

2.3 Participants

A total of sixteen human subjects participated in this study (mean age= 40.6 years old). Each participant was a licensed daily driver without any vision limitations or disabilities. The participants were divided into four groups according to their gender (male and female) and age (young and old). The old group represents drivers over 40 years of age and the younger drivers were from 18 to 35 years of age.

2.4 Experiment Design

As independent variables, road conditions (city roads/highway) and presence of the power-flow gauge (with/without) were manipulated in forms of a factorial design. Thus, each participant was asked to drive the simulator for four trials, consisted of two driving scenarios (road conditions) across other two conditions, the presence of the gauge. In test trials for city road driving, participants were asked to drive an urban road that required the participant to make frequent stops and turning, which were not required in the highway-driving scenario. The speed limits for city highway driving were 40km/h and 60km/h, respectively. The order of trials for participants were randomized and balanced across participants and trial conditions. Before each participant begun the test trials, practice trials were provided in order for them to be familiar with the driving simulator and tasks. During the practice trials, participants were given a demonstration of how the power-flow gauge works. They were told that their aim was to drive the vehicle normally while trying to keep the power-flow gauge bars towards the left, and green. In addition to experimental conditions, the effects of driver profiles (age and gender) were also observed.

Levels of acceleration and brake pedal pressing positions were recorded during experimental trials as response measures. Then the mean values and kurtosis of the both pedals press positions were calculated and analyzed. In particular, the kurtosis of the pedal pressings is analogous to the standard deviation of pedal position over time. This indicates that lower kurtosis represents less stable pedal pressing behaviors due to making frequent and abrupt corrections in pedal pressing.

2.5 Hypotheses

Based on this study's experimental plan and findings in previous studies, we expected the present of the power-flow gauge interface would decrease means of accelerator and brake pedals pressing positions (Hypothesis (H) 1) as well as increase of kurtosis of the pedal pressing positions (H2) due to stable driving. Along with this, the road conditions were expected to affect pedal behaviors. That is, highway driving would yield less mean (H3) and higher kurtosis (H4) for accelerator and brake pedal pressings compared to city driving.

3 Results

3.1 Effect on Acceleration Pedal Behaviors

Mean of accelerator pedal position: Analysis of Variance (ANOVA) results revealed no significant main and interaction effects of road conditions, presence of gauge, age and gender on mean value of acceleration pedal pressing positions during driving. However, an interaction effect of gender and age was marginally significant ($F(1,38)=3.69$, $p=0.0624$), indicating that the mean of acceleration pedal value for

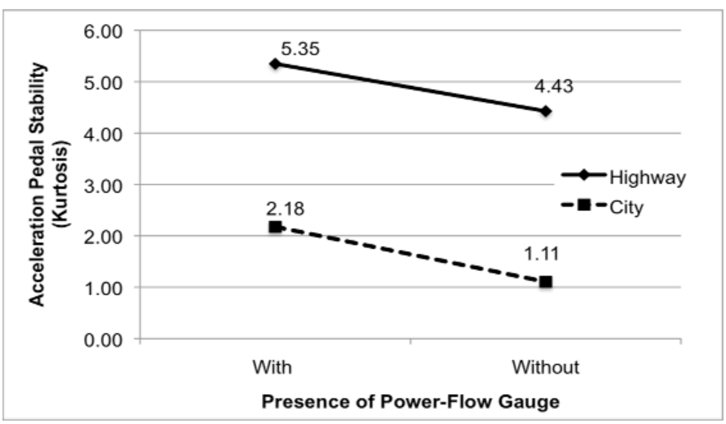


Fig. 3. Acceleration pedal stability (kurtosis) for road conditions and presence of a power-flow gauge

young males was lower than young female group while it is higher in old male than old female. Even though it was not significant, the data and trend shows that the mean acceleration pedal value is lower when driving with the gauge than without power-flow gauge for gender group, age group, and road condition.

Kurtosis of acceleration pedal position: ANOVA results revealed significant main effects of presence of power-flow gauge ($F(1,38)$, $p=0.0329$) and road conditions ($F(1,38)=52.08$, $p<0.0001$) on acceleration pedal behavior (see Fig. 3). Since higher kurtosis (stability in control) indicates higher stable driving, it can be inferred that the drivers were inclined towards more stable (or mild) acceleration pedal pressing when the gauge interface was presented and when they are driving on highway roads.

3.2 Effect on Brake Pedal Behaviors

Mean of brake pedal position: There were no significant main effects of road conditions and presence of the power-flow gauge on mean of brake pedal pressing. However, ANOVA revealed the effect of age ($F(1,38)=11.86$, $p=0.0014$) and the interaction effect of age and gender ($F(1,38)=20.16$, $p<0.0001$) to be significant on the mean brake value. The young participant group had a greater mean brake value than the older ones. The young males had a higher mean brake value than the older males or females. The interaction between gender and the presence of power-flow gauge was marginally significant ($F(1,38)=3.51$, $p=0.0686$), along with the interaction between age and the presence/absence of power-flow gauge ($F(1,38)=3.49$, $p=0.0695$). The males were found to have higher mean brake value with the power-flow gauge as compared to, without. The females were found to have higher mean brake values without the power-flow gauge as compared to with. Also, the young subjects were found to have a higher mean brake value than the older ones. This could be as a result of the slow reaction times of the older subjects.

Kurtosis of brake pedal position: ANOVA results did not reveal the effect of the existence of the power-flow gauge on stability with regards to brake pedal pressings. However, as expected, the braking behavior was significantly more stable (higher kurtosis) for highway driving than for city driving ($F(1,38)=7.32$, $p=0.0102$). This may attributable that there were fewer stop and go situations in highway driving than in city driving. In addition to this, ANOVA results revealed the effect of age on brake pedal pressing stability to be significant ($F(1,38)=4.11$, $p=0.0496$). The young subjects were found to have a relatively stable braking pattern as compared to the older subjects.

4 Discussion and Conclusion

In this study, the effects of a power-flow gauge interface for electric vehicles on drivers' acceleration and braking behaviors were examined by an experiment. While the mean values of accelerator and brake pedal pressings were not affected by the present of the power-flow gauge interface (not in-line with H1), the experiment results demonstrated the stability of the accelerator pedal pressing was increased when using the interface, which is in-line with H2. With a similar pattern, the results for effects of road conditions were not in-line with H3 (less means in highway driving for both pedals) but confirmed that higher stabilities for acceleration/brake pedal behaviors in highway driving (H4). Since each participant was asked to drive the simulation with specific driving speed on each driving condition, it can be inferred that the presence of power-flow gauge interface encouraged the drivers to maintain a relatively steady acceleration as well as discouraged them from being too aggressive while driving. This supports findings in previous studies that the use of the interface could affect driving behaviors. It also may indicate that the changes in driving behavior may affect energy consumptions in forms of economic driving, not limited on gas vehicles.

The study has several caveats, including limited driving performance measures due to the use of a relatively low-fidelity simulator. In addition, the exact electric consumption during driving was not measured in the experiment since the formula for calculating power consumption may be affected many factors such as motor/vehicle dynamics, which are varied for different vehicles and driving contexts.

In regard to future research, it would be interesting to investigate an optimal feature of power-flow interfaces with various manipulations of interface features (e.g., shapes and locations) as well as measurements of driving performance (e.g., speed, lane keeping, detailed dynamic energy consumption, and driver distraction using an eye tracking system).

References

1. The Alliance of Automobile Manufacturers, `http://www.ecodrivingusa.com`
2. Evans, L.: Driver behavior effects on fuel consumption in urban driving. Human Factors 21(4), 389–398 (1979)
3. Waters, M.H.L., Laker, I.B.: Research on fuel conservation for cars. Report No. 921. Transport Research Laboratory, Crowthorne, England (1980)
4. Larsson, H., Ericsson, E.: The effects of an acceleration advisory tool in vehicles for reduced fuel consumption and emissions. Transportation Research Part D 14, 141–146 (2009)
5. Voort, M., Dougherty, M.S., Maarseveen, M.: A prototype fuel-efficiency support tool. Transaction Research Part C 9, 279–296 (2001)
6. Jenness, J.W., Singer, J., Walrath, J., Lubar, E.: Fuel economy driver interfaces: Design ranger and driver opinions. Task 1 and Task 2 Report, DOT HS 811 092, Washington, DC: National Highway Traffic Safety Administration (2009)
7. Manser, M.P., Rakauskas, M., Graving, J., Jenness, J.W.: Fuel Economy Driver Interfaces: Develop Interface Recommendations. Report on Task 3, DOT HS 811 319), Washington, DC: National Highway Traffic Safety Administration (2010)
8. A Glass of Water, `http://www.aglassofwater.org`

In-Car Dictation and Driver's Distraction: A Case Study

Martin Labský[1], Tomáš Macek[1], Jan Kleindienst[1],
Holger Quast[2], and Christophe Couvreur[2]

[1] IBM Czech Republic, V Parku 2294/4, 148 00 Praha 4 Chodov
{tomas_macek,martin.labsky,jankle}@cz.ibm.com
[2] Nuance Communications, Guldensporenpark Building D, Merelbeke, 9820 Belgium
{holger.quast,christophe.couvreur}@nuance.com

Abstract. We describe a prototype dictation UI for use in cars and evaluate it by measuring (1) driver's distraction, (2) task completion time, and (3) task completion quality. We use a simulated lane change test (LCT) to assess driving quality while using the prototype, while texting using a cell phone and when just driving. The prototype was used in two modes – with and without a display (eyes-free). Several statistics were collected from the reference and distracted driving LCT trips for a group of 11 test subjects. These statistics include driver's mean deviation from ideal path, the standard deviation of driver's lateral position on the road, reaction times and the amount and quality of entered text. We confirm that driving performance was significantly better when using a speech enabled UI compared to texting using a cell phone. Interestingly, we measured a significant improvement in driving quality when the same dictation prototype was used in eyes-free mode.

1 Introduction

The popularity of using various communication, navigation, entertainment and driver assistance systems in cars is steadily increasing as more and more of these systems enter the market at competitive prices. Text entry is one domain not covered by current production systems. According to [4], about 30% of drivers surveyed in Australia sometimes entered a text message using their mobile phone while driving and 1 out of 6 drivers did so regularly. At the same time, receiving and especially sending messages is perceived by drivers as one of the most distracting tasks [11]. As a consequence, the number of distraction-related crashes is thought to be increasing. The primary constraint when developing automotive UIs is thus to keep driver's distraction minimal. The secondary aims are to minimize task completion time and maximize task completion quality. In this paper, we evaluate a prototype text dictation UI according to these constraints and aims.

2 Related Work

A significant amount of research has been done that compares driving performance degradation due to using conventional and speech-enabled UIs [1]. The general conclusion is that while speech UIs still impact driving quality, they do so significantly

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 418–425, 2011.

less than conventional UIs. Most distraction caused by conventional systems seems to be due to drivers looking away from the road, as measured by the number and duration of eye gazes. In addition, using speech was observed to be faster for most evaluated tasks.

A number of approaches were also described to perform dictation in hands-busy environments [9]. In particular, hands-free text navigation and error correction were addressed by [8]. The impact of the most prevalent correction method, re-speaking, was evaluated by [10]. Microsoft described a prototype system [5] that allowed responding to incoming text messages by matching message templates.

3 Experimental Design

The dictation prototype evaluated in this paper, code-named *ECOR* (as for error-correction), allows for open domain, unconstrained dictation augmented by a number of methods that support the tasks of error detection, location and correction. While the system has a fully-fledged GUI, it can be used completely eyes-free in order to minimize driver's distraction. The system echoes recognized text chunks as they are dictated (using clearly speaking TTS) and allows for navigating and correcting dictated text.

A standard LCT simulator [6] was used to simulate driving in an office environment. The simulator was shown on a 22" screen and the *ECOR* screen showed on a separate 8", 800x600 touch-screen, positioned on the right side of the simulator screen. A Logitech MOMO steering wheel and pedals were used to control the simulator and 4 buttons (incl. push-to-talk) on the steering wheel controlled the prototype. Our setup was very similar to that used by [3,7] and to that described as "PC condition" by [2].

One LCT trip consisted of a 3km straight 3-lane road with 18 irregularly distributed change lane signs (lane changes of both 1 and 2 lanes were included). The evaluated segment started with an extra "START" sign and ended 50m after the last change lane sign. Drivers kept a fixed maximum speed of 60km/h (16.7m/s) during the whole trip.

4 Evaluation

A group of 11 test subjects (9 male, 2 female, aged 19-43, mean 33) was used to measure the degradation in driving performance due to distraction caused by entering text.

4.1 Procedure

First, each subject was allowed to train driving until they mastered the LCT. Then, 2 **reference** LCT trips were collected. One was used to compute an adapted model of the driver's ideal path and the second was used to compute driving performance statistics using the adapted ideal path.

The **ideal path** was modeled using a linear poly-line. Because each driver had a slightly different driving style, we adapted several parameters of the ideal path to accommodate for individual driving styles, in order to make the driving performance statistics more comparable among drivers. The major differences among drivers during their reference drives are listed below along with the corresponding adapted parameters of the ideal path:

- Different steering angles and durations of lane changes (maneuver lengths when changing 1 or 2 lanes in both directions).
- Different reaction times to lane change signs (distance before the lane change sign where the maneuver starts).
- Different standard driving positions within each of the 3 lanes (lateral car position offset for each lane).

After the reference LCT trips, each subject was introduced to the *ECOR* prototype with display and had sufficient time to practice dictating arbitrary text; initially when car was parked, then also while driving. The time spent practicing *ECOR* was 20-30 minutes, including 1-2 training LCT trips. After mastering *ECOR*, each subject conducted a single LCT trip **with display**, during which s/he was instructed to enter a sequence of text messages with pre-defined semantic content (e.g. "instruct your partner to buy oranges, wine and chocolate" or "tell your secretary to set up a meeting, at the library, tomorrow, at 5pm.").

ECOR display was then switched off and subjects were allowed up to 20 minutes to practice using *ECOR* eyes-free, including 1-2 training LCT trips. This was followed by a single evaluated **eyes-free** LCT trip, during which the same sequence of messages was to be entered.

Finally, the subjects ware allowed 1-2 training LCT trips to practice driving while entering text using their own cell phone. After this, a single **cell phone** LCT trip was conducted while entering messages according to the same specifications. Two out of 11 subjects said they sometimes sent text messages using their cell phone while driving.

Fig. 1. Average MDev and SDLP with 95% confidence intervals

4.2 Driving Performance Results

Driving performance was measured for all 4 evaluated LCT trips using the following statistics.

- The car's mean deviation (MDev) from the ideal path in meters. This measures how much, on average, the driver drove off his/her ideal track, and is computed simply as the absolute value of subtracting the actual and ideal lateral car positions at each sampled point, and by averaging over the distance of the trip.
- The standard deviation of lateral position (SDLP) of the car in meters. SDLP measures how much the driver "weaves" within the lane and is computed as the standard deviation of the car's absolute actual deviation from its ideal path at each sampled point.
- The number of missed lane change signs per trip.
- The number of accidents during which the car's position went out of the road.
- Delay of reaction time to lane change signs.

Mean deviation and SDLP were evaluated over the whole LCT trip, and also only during its lane keeping phases. Averaged values of MDev and SDLP are shown in Figure 1. For the reference trip, both overall MDev and overall SDLP values were around **0.4m**. The next best result was achieved by the eyes-free setup, which had overall MDev and SDLP of **0.51 and 0.54m**, respectively, showing degradation of **10 and 15cm** on average. For the display trip, distraction was higher, with overall Mdev and SDLP of **0.73 and 0.93m**, with average degradation of **32 and 56cm**. The worst results were measured for the cell phone trip with overall MDev and SDLP of **1.11 and 1.15m**, with degradation of **70 and 77cm**. For the lane-keeping phases, we naturally observed lower MDev values, starting at Reference **0.32m**, followed by **0.37, 0.51 and 0.96m** for the Eyes-free, Display and Cell phone trips.

It is important to note that adapting the ideal path dramatically reduced abs. values of both MDev and SDLP, as opposed to using a single predetermined path. In our case, using a default unadapted ideal path would cause the absolute values of Mdev to be about **1m** higher for all types of trips, and up to **1.5m** higher for SDLP. Therefore, comparing ratios of these statistics for distracted and undistracted LCT trips across studies can be misleading.

Per-subject results (for 11 subjects) for overall MDev and overall SDLP for all 4 LCT trips are shown in Figure 2. We can see that the order, by level of distraction, is for most subjects as follows: Reference < Eyes-free < Display < Cell phone. We can also see that two subjects had SDLP significantly greater for the *ECOR* Display trip than for the Cell phone trip – in both cases this was due to the subject missing a sign in the *ECOR* trip but not missing any sign in the Cell phone trip[1].

---

[1] The impact of missing a lane change sign is dramatic both to the overall MDev and overall SDLP, since the actual path diverges from the ideal path by the width of 1 or 2 lanes for the whole lane-keeping range until the next sign. When computing the lane-keeping versions of MDev and SDLP in Figure 1, these impacted segments were excluded.

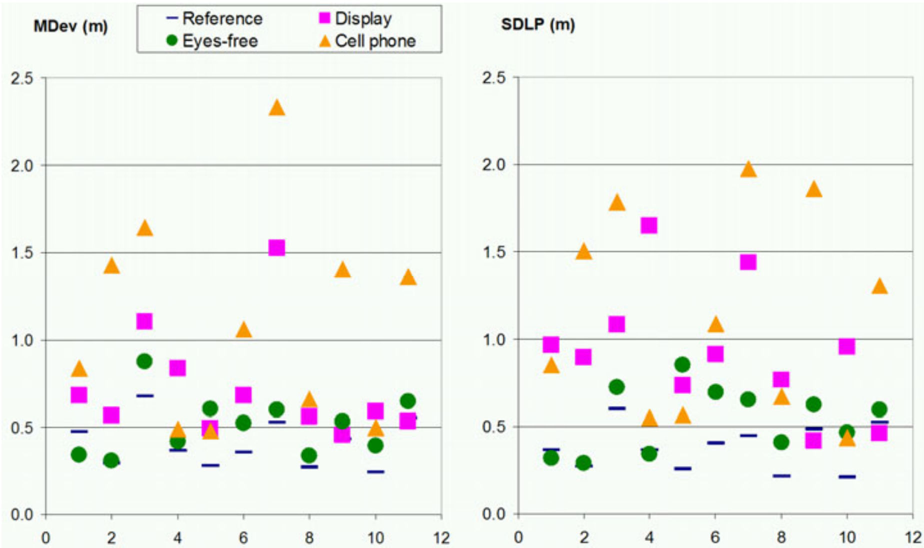


Fig. 2. Detailed values of MDev and SDLP for 11 subjects

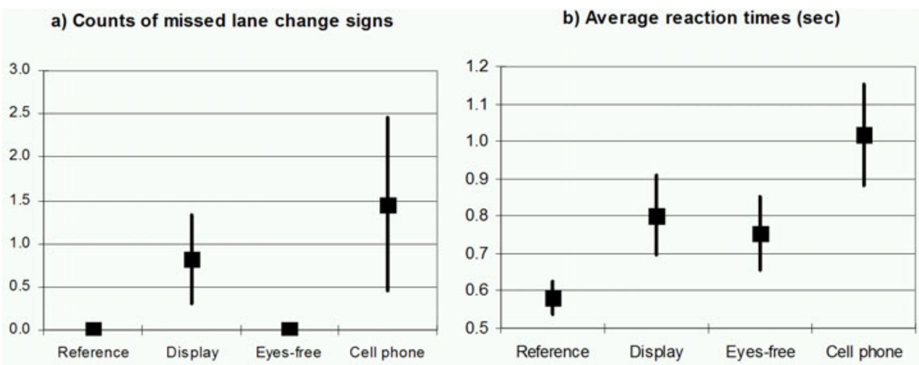


Fig. 3. Average numbers of missed lane change signs and average reaction times across all 4 LCT trips, all with 95% confidence intervals

Apart from MDev and SDLP, we also measured the number of **missed change lane signs**, computed by visual inspection of the actual and reference tracks using the LCT analysis tool; see Figure 3a. Missed signs only occurred when the secondary task included visual distraction and the number of missed signs had a high variance. The averages were **0.8 missed signs** per *ECOR* trip with display and **1.45 missed signs** per Cell phone trip. **Out-of-the road accidents** were all events when the actual car path went out of the road. These occurred only when operating cell phones (**0.7 accidents per trip**) and had a high variance – one subject had 4 accidents, another had 2, and two subjects had 1 accident each.

Driver's reaction times to lane change signs were measured as follows: the actual car path was examined in the range between 35m before the sign (sign visibility) and 20m after the sign. The first noticeable steering wheel movement of angle greater than 3° was identified and its start was considered as the time of driver's reaction to the sign. Reaction times shown in Figures 3a and 4 were computed by subtracting the time the sign became visible from the time of driver's reaction. When undistracted, drivers on average reacted in **0.58s**. The next best reaction times were **0.75s** for the Eyes-free setup and **0.8s** for the system with display. Cell phone users reacted in **1s**. Figure 4 shows 198 lane change reactions (11 subjects × 18 signs) for each of the 4 LCT trips. Note that Eyes-free users had less extreme delays in their individual reactions.

Finally, we analyzed subject's eye gazes during the 3 distracted LCT trips by manually annotating their videos. Table 1 shows highest road attention for the Eyes-free drivers. Subjects spent **9%** of their driving time looking at the display when it was available. For cell phone, the average time spent looking at its display was extreme **44%**. As cell phone typing consisted of many elementary operations, the number of gazes at its UI was also 3 times higher compared to speech interface with display.

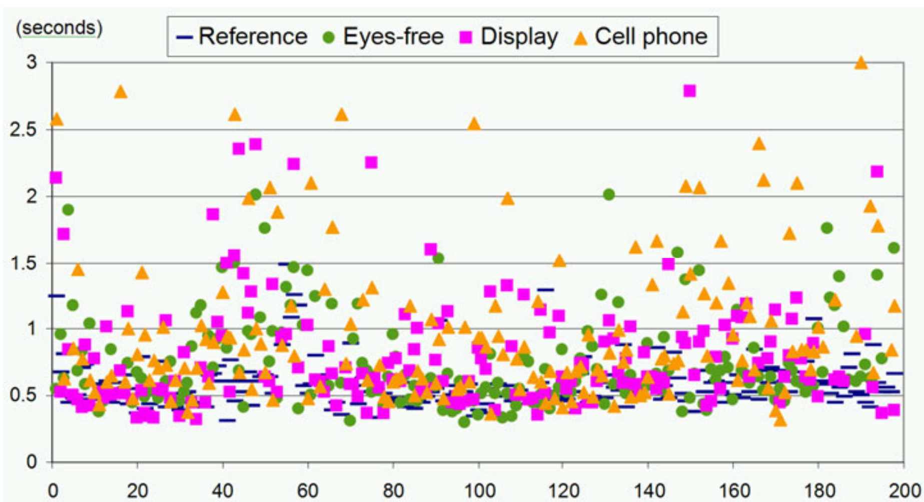


Fig. 4. Per-sign reaction times for all 4 LCT trips

4.3 Dictation Performance Results

In addition to driving performance, we also evaluated the subjects' performance of composing messages. For all 3 distracted trips, we collected texts of all composed messages. These texts were analyzed manually by a single test conductor and each message was scored on a 0-1 scale. High scores were assigned to messages that contained all prescribed semantic content and did not contain undesired text. The major scoring factor was the semantic understandability of the message. Typos that could be easily decoded had minimal impact on the scores.

In Figure 5 we can see that using voice to dictate messages was on average **significantly faster** than typing using mobile phone. On the other hand, **message quality was lower** for messages entered by voice. This is due to ASR errors being more destructive to message semantics as they typically mistake one or more whole words for other valid words. If not noticed and corrected by the user, ASR errors cause semantic mismatches that are much harder to decode by the addressee than isolated character typos often produced when typing messages manually.

When comparing the *ECOR* tests with and without display, message quality was slightly worse for the messages entered without display. Surprisingly, subjects were able to send slightly more messages without display, which we attribute to the subjects not noticing some recognition errors and therefore getting less “stuck” correcting dictated text.

4.4 Discussion

According to MDev, SDLP, missed sign counts and reaction times, the eyes-free dictation system resulted in significantly better results than the same system with display switched on. Based on the gaze data, we conclude that the visual distraction caused by reading text appearing on the display has significant adverse effects on driving. When available, subjects looked at the display even though all information was available via acoustic feedback.

Unlike [5], we consider the eyes-free use of an automotive dictation system possible by solving the problems of error detection and correction. Several features aiding both problems were already part of the evaluated prototype and several more are yet to be evaluated.

Table 1. Portions of time spent looking at the road, at the text entry UI, and elsewhere. Average counts and durations in seconds are shown for out-of-the-road gazes.

System \ Gazes at	Road	Text UI (%, #, sec)			Other (%, #, sec)		
Display	85%	9%	19	0.74	6%	23	0.46
Eyes-free	94%	-	-	-	6%	20	0.47
Cell phone	54%	44%	59	1.31	2%	4	0.70

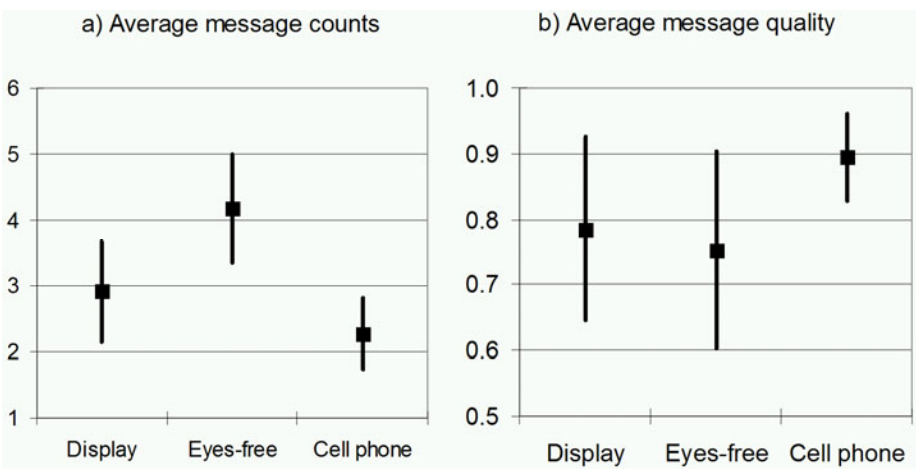


Fig. 5. Numbers of sent messages and message quality

5 Conclusion and Future Work

We presented results of evaluating an automotive text dictation system using a LCT simulator. We showed that when operating without a display, driving performance was significantly better compared to the same system with display. Both versions of the dictation system greatly outperformed cell phone typing in all aspects except for message quality, which we attribute to the character of ASR errors.

Acknowledgements. We would like to thank to our test subjects who participated in the evaluation (honored by hand-made wooden sticks allowing them to use the degree "Subject").

References

1. Barón, A., Green, P.: Safety and Usability of Speech Interfaces for In-Vehicle Tasks while Driving: A Brief Literature Review. Technical Report UMTRI-2006-5. University of Michigan Transportation Research Institute (2006)
2. Bruyas, M.P., Brusque, C., Auriault, A., Tattegrain, H., Aillerie, I., Duraz, M.: Impairment of Lane Change Performance due to Distraction: Effect of Experimental Contexts. In: Proc. European Conf. on Human Centered Design for Intelligent Transport Systems (2008)
3. Harbluk, J.L., Mitroi, J.S., Burns, P.C.: Three Navigation Systems with Three Tasks: Using the Lane-change test to Assess Distraction Demand. In: Proc. 5th Intl. Driving Symp. on Human Factors in Driver Assessment, Training and Vehicle Design (2009)
4. Hosking, S., Young, K., Regan, M.: The effects of text messaging on young novice driver performance. In: Faulks, I.J., et al. (eds.) Distracted Driving, pp. 155–187. Australasian College of Road Safety, NSW (2007)
5. Ju, Y.C., Paek, T.: A Voice Search Approach to Replying to SMS Messages in Automobiles. International Speech Communication Association (2009)
6. Mattes, S.: The Lane-Change-Task as a Tool for Driver Distraction Evaluation. In: Proc. Annual Spring Conference of the GFA/ISOES (2003)
7. Petzoldt, T., Bär, N., Krems, J.F.: Gender Effects on Lane Change Test Performance. In: Proc. 5th Intl. Driving Symp. on Human Factors in Driver Assessment, Training and Vehicle Design (2009)
8. Suhm, B., Myers, B., Waibel, A.: Multimodal error correction for speech user interfaces. In: Proc. ACM Transactions on Computer-Human Interaction, TOCHI (2001)
9. Oviatt, S., et al.: Designing the user interface for multimodal speech and pen-based gesture applications: State-of-the-art systems and future research directions. HCI 15(4), 263–322 (2000)
10. Vertanen, K.: Speech and Speech Recognition during Dictation Corrections. In: Proc. Interspeech (2006)
11. Young, K.L., Regan, M., Hammer, M.: Driver Distraction: A Review of the Literature. Monash University Accident Research Centre. Report no. 206 (2003)

Driver's Experience and Behavioral Patterns through the Observation of Commercial Vehicle Driving

Youngjae Lim [1], Sungjoon Park [2], Eui S. Jung [1], and Taeil Kim [1]

[1] Department of Industrial & Management Engineering, Korea University,
5-1 Anam-dong, Seongbuk-gu, Seoul, Korea
{ergoim,ejung,ergokim}@korea.ac.kr,
[2] Department of Industrial & Management Engineering, Namseoul University,
21Maeju-ri, Seonghwan-eup, Seobuk-gu, Cheonan, Korea
{sjpark}@nsu.ac.kr

Abstract. The objective of this paper was to investigate the driver's experience of commercial vehicle driving. This activity analysis was performed to account for anything that a person does in the cabin. The main user activities were classified into three groups; critical, significant and moderate. The user activities were also analyzed with respect to ergonomic approaches such as using an ergonomic checklist and an expert evaluation. The evaluation results showed that there was not enough space among the seat control buttons. In addition, the center console lacked storage facilities compared to the occupied space. The observation of behavioral patterns was conducted to guide an observer though the user environment. A total of 11 commercial vehicle drivers were recruited for this study. The 17 representative behavioral patterns were defined in terms of the frequency and severity of events. By taking advantage of observations, a variety of recommendations could be made from the user-centric perspective.

Keywords: User Experience, Observation, Behavioral patterns.

1 Introduction

In the past, the evaluation of products by users has only been confined to matters of functionality or quality. As levels of income increase and cultural development continues, the requirements of users will expand to more diverse areas. This means that the product development concept shifts from a manufacturer-active paradigm to a customer-active paradigm, one that should comply with a user's statement of what they want [1]. Recently, customers have become more involved in product development and the industry has been gradually promoting the efforts to reflect the user experience. Generally, the idea of the user experience means all activities representing the various interactions between a user and a product [2, 3]. Roto [4] mentioned a comprehensive user experience that not only built up momentary pleasure but also affected continuing changes in attitudes and feelings for the product.

In the automotive industry, the user can define a vehicle in a variety of ways, from simple transportation to essential living area or even mobile office space [5].

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 426–434, 2011.

Determining the ideal customer requirements and providing a pleasant and comfortable space are increasingly necessary. Thus, the research scope of automotive ergonomics should be expanded to reflect a variety of customer activities that occur in real-life situations as well as an enhancement of affection in qualities related to the interior package. The overall index of perceived quality is actually dependant on the interior package design with a high concentration of interface [6].

Several studies on interior packages of sedans and SUVs (Sports Utility Vehicles) were performed for the development of an easy-to-use and human-centric vehicle. Herriotts [7] conducted an extensive nation-wide survey to produce clear guidelines for vehicles that were easy to get in and out of by older drivers. Wang [8] aimed to evaluate the comfort of automobile clutch pedal operation and possible correlation between subjective rating and biomechanical parameters. However, the research on commercial vehicles dealt mainly with driving performance in regard to fuel efficiency. Loading of goods and overall durability were also areas of interest to researchers. The driver's expectation of the interior package of a commercial vehicle was strongly influenced by the vehicle's price. Unfortunately, the ergonomic improvement was relatively insufficient. One consistent feature described by drivers of commercial vehicles was the long hours they spent away from home. For this reason, the portion of residence area in the vehicle cabin was usually greater than other segment.

To improve the user experience of commercial vehicle drivers, suggestions centering on an ergonomic seat or innovative console design seem very important. The customer survey or interview is one of the conventional techniques to extract driver feedback. However, techniques like these are often influenced by the memory or intention of respondents and make capturing an accurate account of the overall experience somewhat difficult. An on-site observation has proven much more useful in recognizing potential needs and behavioral patterns in recent years.

This study begins by constructing user activities that can be experienced during the act of driving a commercial vehicle. Secondly, methods such as using an ergonomic checklist and other expert evaluations were carried out to analyze the interaction between the driver and the package components concerned with major user activities. Then, the actual behavior of commercial vehicle drivers was observed directly and patterns based on driving conditions were investigated. As a result, issues surrounding physical conditions, mental models and interior packages were suggested in this study. This research observes user activities and driver experiences in real-life situations. It is expected that ergonomic design based on the experience and behavioral characteristics of drivers will enhance user satisfaction for commercial vehicles.

2 Methods

2.1 Analysis of the User Experience

Commercial vehicle drivers eagerly require an automobile that fits their lifestyle as much as it does the demands of their profession. To achieve this objective, an analysis was carried out to determine what activities are critical to user experience in the vehicle cabin. This study performed a brainstorming session on the topic of driver activity and the eight participants were a mixture of graduate students majoring in ergonomics and practical professionals. Findings included driving activities and non-driving situations like resting, eating in the rear seat, etc. There were 355 general activities based on user experiences that were extracted with respect to the package design.

Overall, the aforementioned activities were too broad to apply to the interior package of a commercial vehicle. Thus only those activities specifically related to the research objective were selected. The activities that were not relevant to a commercial vehicle (e.g., identifying a bag on the passenger seat) were excluded. In addition, various related activities (e.g., removing/replacing a beverage container from/in the cup holder) were integrated with one another for the purposes of the study. In the end, a total of 59 user experiences relevant to commercial vehicle drivers were selected.

The frequency of use and discomfort of the user experience were also investigated in order to identify priorities and preferences. A survey measuring each activity was conducted among 15 commercial vehicle drivers using a five-point (1 – "rarely used" to 5 – "frequently used") Likert scale.

2.2 Interactive Evaluation of the Interior Package

An analysis of package items with respect to a driver's behavioral pattern and cognitive structure was needed. This would allow for more emphasis to be placed on the user experience with regard to complicated functions. This study suggested ergonomic evaluation methods to assess a variety of design factors.

Ergonomic checklist. A human characteristic was defined as a factor that indicated an ability or limitation of a driver's physical or mental status. Buner and Kumar [9] mentioned that an analysis of human characteristics was required to identify specific reasons for a preferred product. Human characteristics were classified according to the stage of cognition and response: sensing, information processing, motor function and anthropometry. Also, this research determined the evaluation criteria based on the main activities and package items of a commercial vehicle. Selected criteria were matched with the ergonomic design principle and transformed into a checklist. Therefore, an in-depth analysis was available for the existing package items.

Expert evaluation. An expert evaluation is a usability inspection method for development that helps to identify usability problems [10]. It specifically involves professionals who examine the interior package and judge its compliance with a recognized usability principle. This evaluation method is now widely practiced in the automotive ergonomics sector. Furthermore, this method can be very efficient if time or budget restrictions are in effect. In this study, the ergonomics experts qualitatively analyzed elements of usability issues related to the package. Issues such as controlling interior switches, accessing residence areas and having acceptable clearance for indoor activities were all assessed through a group discussion.

2.3 Participatory Observation

The user observation method involves firsthand accounts of individuals who go into an actual situation and share their experiences [11]. The behavioral observation is an appropriate method to capture important scenes by recording the activities of product users. In order to observe detailed behavior, one or more cameras recording various angles are needed in the cabin. However, it is desirable that the driver performs his natural behavioral pattern and does not let the video cameras affect his normal activities.

A total of 11 commercial vehicle drivers were recruited for this study. The experimenter obtained consent for the research objective in advance and then explained the recording procedure. To monitor the emergent situations or the activities away from the fixed camera angles, researchers would ride along in the cabin and use a handheld camcorder. Also, the observation of behavioral characteristics while driving at night was facilitated using an infrared camera.

3 Results

3.1 The Grouping of User Activities

The main user activities were classified by priority according to their frequency of use and discomfort. Analysis of variance was conducted on the collected data using the statistical analysis system. Then the post-hoc analysis, known as Student–Newman–Keuls (SNK), was carried out to derive statistically similar groups. As a result, the 59 user activities were categorized into three groups (critical, significant and moderate) based on the requirement level, and the total number of activities for each group was 3, 9 and 13, respectively. Table 1 illustrates a sampling of the results.

Table 1. The results for defining the requirement level of user activity

Requirement Level	Activity
Critical	Identifying a gauge-cluster
	Operating a navigation system and viewing the DMB
Significant	Throwing away tobacco into an ash-tray
	Putting a water bottle into a cup-holder
	Moving between the driver's seat and rear-bed
Moderate	Operating an audio control on center-fascia
	Adjusting the brightness of room light
	Making a phone call using the Bluetooth
	Pulling out a ticket at a tollgate
	Opening (or closing) the sunroof

3.2 Outcomes of Package Evaluation

Analysis of the ergonomic checklist. The ergonomic evaluation criteria were created with respect to vehicle features and packaging elements. Then a checklist analysis was carried out for elements such as seat adjustment, lumbar support control, cruise control, dome light, etc. In the case of lumbar support control, a total of six human characteristics were identified in the evaluation criteria. The related design principles were evaluated through assessment questions and some of the detailed examples are illustrated in Table 2.

Table 2. The checklist for evaluation of lumbar support control

Evaluation Criteria	Ergonomic Checklist
Affordance	Did the shape of control panel enable blind operation?
	Was it designed as a lever type (manual) or a push button type (automatic)?
Clearance	Were controls sufficient in the allotted space (7.5cm, 5cm above the horizontal and vertical)?
Compatibility	Did the direction of control movement coincide with that of the lumbar support?
Reach	Was it close enough without bending the driver's waist or twisting his trunk?
Symbol Coding	Did the symbol associated with the direction suggested reduce confusion?
Visibility	Was the location of control easy to find?
	Did the lights installed make it easier to look at the location of the controls?

The evaluation results showed that there was not enough space among the control buttons in terms of clearance. As a result, it was noted that the situation could lead to a malfunction. In terms of reach evaluation, control buttons that were positioned on the back of the seat caused interference with other package items. The driver had to twist his shoulder under these circumstances. Furthermore, identifying the location of the control was difficult in diminished visibility situations while driving at night.

Findings through expert evaluations. An evaluation team consisting of four ergonomic professionals visited a commercial vehicle production plant. The experts actually sat in the vehicle cabin to experience the user activity. The group discussed multiple situations and diagnosed some of the package components. Ergonomic considerations were derived from the discussion.

Steering wheel

Center console

Fig. 1. Examples of expert evaluation

The gauge cluster, which was rated in the critical group for requirement level, was identified as a major problem area because glare negatively affected visibility. The tilting variation of the steering wheel in an existing vehicle appeared to be insufficient for optimal use as shown in Figure 1. Improving the structure of the steering wheel would allow more room for ingress or egress, as well as support office activity. Moreover, the center console lacked storage facilities compared to the occupied space. Commercial vehicle drivers expressed a need for increased carrying capacity

Other considerations included the safety handle, which was distributed in several places and seemed to have an uncomfortable handgrip. An excessive wrist bending was observed when drivers applied the center console parking brake. It was proposed that the parking brake be moved to the center-fascia and redesigned using a button-type operation.

3.3 Behavioral Patterns of Commercial Truck Drivers

The total travel distance and recording time of observations were 1200 km and 50 hours, respectively. Through behavioral analysis of driving and non-driving conditions, the 17 representative behavioral patterns were defined in terms of the frequency and severity of events. Table 3 presents a sampling of the results.

Table 3. The representative behavioral patterns

Observed Condition	Behavioral Characteristics
Driving	Using a mobile phone while driving
	Operating a navigation system
	Drinking a beverages and putting it into a cup-holder
	Turning on (or off) an interior light
	Controlling the parking brake
	Folding (or unfolding) a front sun-visor
Non-driving	Stretching of the upper body in the cabin
	Moving between the driver's seat and the rear bed
	Taking a break and adjusting the driver's seat
	Cleaning up the rear-space while the driver is in the front seat
	Taking off shoes (and keeping them off) after a driver rides in the cabin
	Identifying the cargo while waiting for loading

Among the major behavioral patterns, it was noted that the driver's waist was bent too much when using the navigation system because the device was located too far from the driver's seat. Even cabin dimensions of commercial vehicles are that large, one driver had to use a pen to reach the navigation system interface. At this time the driver's field of view was distracted, thus creating an increased risk of a severe accident.

Fig. 2. Behavioral characteristics during driving situation

In addition, a significant number of drivers drank some type of beverage, especially during their longer trips. However, the number and size of adequate bottle or cup holders was lacking compared to the needs of the drivers. The existing cup holders were less frequently used and not suitable for larger containers. Therefore, some drivers kept water and other beverages in a separate box on the center-console or rear-bed.

When moving from the driver's seat to the rear-bed area during non-driving situations, the center-console and parking-brake caused significant interference and were pointed out as major factors impeding mobility. This moving was only possible under conditions where the drivers kept their backs bent and stepped on the armrest. Therefore, the rear-bed area's frequency of use was diminished during break periods or sleeping time. It was also revealed that drivers sometimes used the rear seat when they stopped in a rest area for a break or after they completed the entire workday. In such cases, adjusting the driver's seat forward allowed for more space in the rear seat area. The control mechanism was a manual type and considerable force was required to fold the seat.

Fig. 3. Behavioral characteristics during non-driving situation

Commercial vehicle drivers seemed to recognize the vehicle cabin as their personal space, like their living room or bedroom. For this reason, keeping the cabin clean was important for many drivers. The activity of taking off shoes was one of the characteristics that supported this notion.

4 Discussion and Conclusions

In the past, users were more satisfied with technologically superior products. But today, products were difficult to success if the designers didn't consider usability and behavioral pattern of users [12]. User-related needs such as convenience, livability, and compatibility were also important in the development of the commercial vehicle. Therefore, an evaluation of the ergonomic approach was required for the design of the interior package.

The activity analysis is used to account for anything that a person does while performing a task. The purpose of an activity analysis is to determine the activities that occur in any context of use, as well as how often the activities are performed [13]. In this study, the activities were verified through research and interviews with users. Scenarios can be developed using a combination of these activities that reflect user experiences and these can be applied to the package design.

Further academic and systematic evaluation should be conducted to improve the reliability of the results. This research derived the evaluation items with respect to human characteristics that specifically referred to posture, clearance, and mental workload. In addition, the ergonomic experts assessed the package usability which was represented by symbol confusion, location and method of control. The observation method can be utilized to guide an observer through the user environment as it is encountered in daily life. It was effective in observing the characteristics especially for office activity and the entertainment features as a dwelling space. Obviously, the rear-seat was used differently than that of a conventional vehicle. It provided space for taking a long rest, reading, and storing items.

Improvements that could be reflected in the new design were discussed. For example, the height of the center-console could be reduced, thus allowing drivers to move more easily to the passenger side or rear-seat area. The installation of an automatic tilting button could ease the difficulties involved when tilting the existing steering wheel. Furthermore, observations revealed that a multi-purpose table would be useful for drivers who need a firm and flat surface to write or eat something in the cabin. By taking advantage of similar observations of the customer experience, a variety of recommendations could be made from the user-centric perspective.

This study investigated the user experience of commercial vehicle driving, and the approaches which evaluated the package elements were proposed. The main user activities were classified into three groups; critical, significant and moderate. Therefore, it identified priorities while considering possible trade-offs. The analysis of package items properly reflected the ergonomic checklist based on a user's ability and physical condition. As part of a direct investigation, the results of expert evaluations were also reflected. These approaches contributed an evaluation of the vehicle from a more objective and user-centered point of view.

The observation of behavioral patterns indicated a need for the expansion of storage space and improved indoor movement. Convenience items such as a refrigerator or a device for stretching were also needed. The user experience was formed by characteristics of the commercial vehicle being used as a residence. In addition, prolonged driving situations impacted behavioral patterns in the user environment. Further research is recommended with regard to drivers' behavioral changes over time, as well as further analysis of the quantitative portion among major activities.

References

1. Foxall, G.R.: User initiated product innovations. Industrial Marketing Management 18(2), 95–104 (1989)
2. Alben, L.: Quality of Experience: defining the criteria for effective interaction design. ACM Interactions 3(3), 11–15 (1996)
3. Desmet, P., Hekkert, P.: Framework for Product Experience. International Journal of Design 1(1), 57–66 (2007)
4. Roto, V.: Web Browsing on Mobile Phones -Characteristics of User Experience, Doctoral dissertation at Helsinki University of Technology, Helsinki, Finland (2006)
5. Lee, H.S.: Current Trends and Future Issues on Automotive Industry. In: Procs. 2007 Fall Conference of the Korean Society of Automotive Engineers, Duksan, Korea (2007)
6. Kang, S.M., Park, S.Y., Park, P.: A Study of Automotive IP design based on the Driver's Cognitive Characteristics. Korean Journal of the Science of Emotion and Sensibility 1(2), 81–91 (1998)
7. Herriotts, P.: Identification of vehicle design requirements for older drivers. Applied Ergonomics 36(3), 255–262 (2005)
8. Wang, X., Breton-Gadegbeku, A.L., Bouzon, L.: Biomechanical evaluation of the comfort of automobile clutch pedal operation. International Journal of Industrial Ergonomics 34(3), 209–221 (2004)
9. Bruner, G.C., Kumar, A.: Explaining Consumer Acceptance of Handheld Internet Devices. Journal of Business Research 58(5), 553–558 (2005)
10. Nielsen, J.: Usability Engineering, pp. 115–148. Academic Press, CA (1993)
11. Lee, J.H.: An investigation of the user research techniques in the user-centered design framework. Journal of Korean Society of Design Science 17(2), 77–86 (2004)
12. Mack, Z., Sharples, S.: The importance of usability in product choice: A mobile phone case study. Ergonomics 52(12), 1514–1528 (2009)
13. Ross, T., Burnett, G.: Evaluation of the human-machine interface for complex system interaction: a case study in vehicle navigation systems. International Journal of Human-Computer Studies 55(4), 661–674 (2001)

Predicting the Effects of Time-Gaps for Adaptive Cruise Control (ACC) on Bus Driver Performance

Brian Tsang-Wei Lin[1] and Sheue-Ling Hwang[2]

[1] University of Michigan Transportation Research Institute, Driver Interface Group, 2901 Baxter Road, Ann Arbor, MI 48109-2150, USA
[2] Department of Industrial Engineering and Engineering Management, National Tsing Hua University, 101, Sec. 2, Guangfu Rd., Hsinchu 30013, Taiwan
btwlin@umich.edu

Abstract. Researchers have a deal of attention to the effects on driver performance when driving with assist systems. This article describes modeling approach to simulate the effects of time-gaps for adaptive cruise control (ACC) on bus driver's performance. A concept model was built with the knowledge of modularization, parameterization, and parallel processing. By running the model, the predictions for the effects of five levels of time-gaps were collected in two measures, mean gap and minimum gap. Predictions from the model were validated by the experiment with a verified fix-based bus driving simulator in authors' previous studies. Through the modeling approach, this research provides a theoretical and accurate way to assess effects of time-gaps. To apply this approach to the evaluation on other driving assist systems (e.g. collision warning systems & navigation systems) is the next topic for authors to work on.

Keywords: Adaptive cruise control (ACC); Driver performance modeling approaches; Time-gaps.

1 Introduction

Driving is an intently operation between human and vehicle. Drivers can control the vehicle directly, give orders indirectly, or only supervise the situation according to information display. Depending on the numbers and complexity of tasks the driver needs to pay attention to, driver's behavior and performance may change. Therefore, driving assist systems are greatly developed for driver's convenience and driving safety (to reduce driver's workload and increase performance). One such helpful system is the adaptive cruise control (ACC), which has been optionally attached to some passenger cars in the market (e.g. Toyota Camry 2010; Lexus 2009, and Mercedes 2008). ACC is the enhanced cruise control that keeps a pre-selected time-gap (gap divided by speed) between the lead and the driver's vehicle. Some studies have evaluated subjective and objective effects of ACC with different measures (Sato, et al., 2006; Suzuki & Nakatsuji, 2003; Törnros, et al., 2002), by experimental results. In author's previous research, time-gap is indicated as the key factor for safety, and proper time-gap settings can lead to better performance and can compensate the in-vehicle distraction (Lin, et al., 2009; Lin, et al., 2008). To predict the effects of time-gap with computational modeling methods is the primary objective of this research.

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 435–443, 2011.

A modeling approach for driving with assist systems should be created carefully. Researchers used to develop various driver behavior models with some modeling methods, such as the driver's action modeling with COSMODRIVE (Bellet & Tattegrain-Veste, 1999), lane-change detection with model tracing methodology (Salvucci, et al., 2007), cellular phone dialing effects with ACT-R (Salvucci & Macuga, 2002), and driver's workload with QN-MHP (Wu & Liu, 2007). This research purposes to realize a conceptual model in the first place and an integrated modeling approach is used to combine and arrange several subtasks into one.

2 Literature Review

2.1 Driver Performance Modeling

Driving is a common but complex task that involves multiple critical subtasks. In order to know how people perform in these tasks, many models have been developed to describe (conceptual) and simulate (computational) driver's behavior. Conceptual models are the foundation of computational ones, and can also help the components and procedure of the driving task be understood such as the Risk-based model (Van der Molen & Bötticher, 1988). As to computational models, two categories are usually included, cognitive and performance models, based on their objectives. Cognitive models are used to simulate the mental statement of human beings, such as perception, memorizing, decision making, anticipating, or option selecting (e.g. COSMODRIVE (Bellet & Tattegrain-Veste, 1999; Tattegrain-Veste, et al., 1996)). These kinds of models are to develop artificial cognitive systems that can represent human's cognitive activities.

Nowadays, some well-developed computational models have both features of cognitive and performance models, that can simulate the whole procedure of the driving task (e.g. ACT-R driver behavior model (Salvucci, 2006), QN-MHP (Tsimhoni & Liu, 2003), SmartAHS (Delorme & Song, 2001)). These models are structured by cognitive models and implemented as performance models. For instance, in the ACT-R driver model, it is based on the ACT-R cognitive architecture (Anderson & Lebiere, 1998), which works with chunks of declarative knowledge and production rules that operate the knowledge. Generally, a human driver model should be put into practice, based on input variables, cognitive processing speed, and how drivers evaluate performance.

2.2 Components in the Model Simulation Cycle

A model simulation intends to generate a sequence of behaviors as humans do, and outputs by these behaviors will give feedback which can activate next stimuli and form a cycle. In the sequence, stimuli will be perceived and comprehended (in cognitive architecture and through memory storage), so that the responses and motions can be properly controlled. How our model simulation works is interpreted in this section.

Stimulus input and perception. The environmental stimuli are from the external world, and models always begin here. In driving models, inputs (stimuli) are generated from the driver's surroundings, by the in-vehicle interface and the outside world. In-vehicle interface takes all the information represented in the vehicle, such as the navigation information (route planning, electronic maps), vehicle conditions (speed, rpm, oil level, engine temperature), multimedia displays (radio, CD player), or communication devices (cellular phone, wireless). The outside world involves the road factors (gradient, curvature, surface, lane numbers), traffic conditions (traffic flow, density), and weather. The input of environmental stimuli will trigger the functions of perceiving stimuli.

Cognitive architecture. In cognitive architecture, knowledge, performance, and learning are included (Anderson & Lebiere, 1998). The knowledge determines what rules should to be triggered with the greatest gain and the performance will generate the actions of the rules. Learning can affect the knowledge and performance by adding new knowledge into memory or practice to enhance existing knowledge. Further, memory is an important component in cognitive architecture. Long-term memory affects the accuracy of perception and the decision making process. Short-term memory is the temporary storage room for what were perceived. In the modeling process, memory continuously interact with the performance to give proper commands to body parts. The performance module has the tactical process and operational process. The tactical process is a complicated part in the performance module. How a driver makes a decision and how he thinks over the current situation are not based on the state of the world, but on the mental representation from perception. From the mental cognition, decisions will be made by rules from the long-term memory, as the ACT-R (Anderson & Lebiere, 1998) and COSMODRIVE (Tattegrain-Veste, et al., 1996) do. These rules provide strategies, priorities, or conditions for (multiple) tasks. The relationship between tactical module and long-term memory is established by driver's knowledge, experience, and skills. Then the outputs of the tactical process are commands for the operational process, to move, control, or respond.

Motor control. The operational module will control the vehicle in two dimensions or operate the in-vehicle task interface. Because different body parts and accuracy are necessary for each task, the action time will also differ. The movement time of some body parts has been observed and estimated according Fitt's Law (Fitts, 1954), such as the one-finger typing movement time (MacKenzie, 2003), using head-controlled computer input devices (Radwin, et al., 1990), and using scrolling and hierarchical lists (Cockburn & Gutwin, 2009). For motions, several critical functions for manual control should be referred and applied in this study. These functions are based on experiments, observations, physiological theories, and HCI theories.

3 Method

In this section, the scenario and environments for drivers were based on author's previous studies (Lin, et al., 2009; Lin, et al., 2008), which was the same scenario for simulation. The concept model was illustrated, coming with the tactical and operational process simulation. Finally, the simulation results were put into comparison with empirical data.

3.1 The Concept Model

To realize a human driver model, in general, there are five common components: environment, perception, memory and rules, tactical process, and operations. The process flows as the following brief (also see Fig. 1). External environment triggers the perception to receive the stimuli from the outside world (Parameterization 1). Then the stimuli will be encoded to meaningful information, which is so-called mental representation (e.g. perceive the brake light of the lead vehicle as a warning signal). The information is temporarily stored in driver's short-term memory (as the factual knowledge) and passed to the tactical process (Parameterization 2). In the tactical process, the decision making is based on the factual knowledge and ruled by the skill knowledge (Parameterization 3), so that the strategy and anticipation can be developed. How to operate, control, and react will be determined and the commands are sent to the operational process (Parameterization 4). All responses and body movements occur here, and will finish the modeling cycle. The learning function, which feed the operational results back to driver's knowledge (Parameterization 5), is involved in the model, but will not be realized in this research.

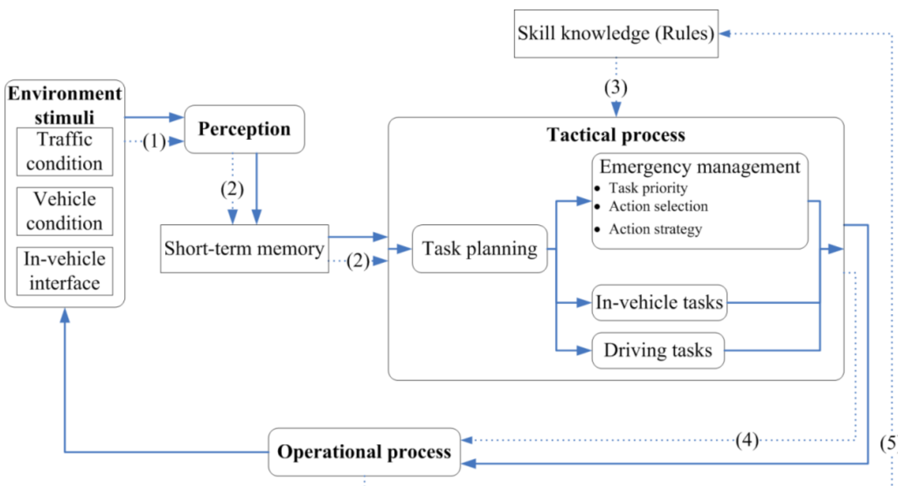


Fig. 1. Structure of the concept driver model. The solid lines represent the flow of the simulation; the dotted lines represent the process of parameterization.

3.2 Tactical and Operational Process Simulation

Comparing to the environment and perception components, memory and tactical process were relatively complicated and affected how the model works greatly. In the component of tactical process, subjects did the mental representation and made decisions for their operations. This model assumed that there were two phases in tactical process, task planning and operation preparation. The former represented how the driver was aware of the current situation and determine what task should be done in this moment. And the latter included handling the general driving tasks, in-vehicle tasks, and the response to emergency situations.

Task planning. In the scenario, drivers were asked to avoid dangers with controlling the brake pedal when the emergency occurred. It was observed that the cue for drivers to determine the reaction for emergency was the current gap. Due to that bus drivers' subjective acceptance for pre-selected time-gaps in the pilot study were the ones over 1.60 s, time-gaps of 1.76 s and above would only be selected to this simulation. It was found out that there was a relationship between the current time-gap which was determined to brake and the pre-selected time-gap settings, as shown in Eq. (1). The polynomial fitting line was shown in Fig. 2, with R^2=.9437. This function was used to simulate that the subject would decide to brake or keep working on general driving and in-vehicle tasks.

$$\text{The current time-gap determined to brake} = -0.1737x^2 + 1.6696x - 0.6902, \quad \text{where } x = \text{pre-selected time-gap (1.76 s to 2.40 s).} \tag{1}$$

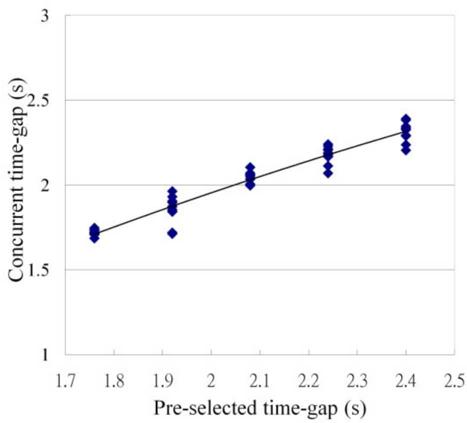


Fig. 2. Function to determine the reaction for emergency

Emergency management. When the driver's emergency management module was triggered, it was assumed that the priority to avoid the danger would be put to the first place at once. After the driver made up his mind to brake, the next step was to set up how the brake would be controlled. The braking movement time, representing the time interval from the brake pedal just moving until completely pushing-down, also has something to do with the pre-selected time-gap, as shown in Eq. (2). The shorter the braking movement time was, the more urgent the situation could be. The polynomial fitting line with R^2 of .9916 was in Fig. 3.

$$\text{The braking movement time} = -0.4018x^2 + 2.1786x - 1.6122, \quad \text{where } x = \text{preselected time-gap (1.76 s to 2.40 s).} \tag{2}$$

Vehicle control. To simulate the primary driving task, it was simple because the driver only needed to steer under the assist of ACC. It is assumed that the action of steering would transfer 2° the most in 100 ms, which was observed and collected from the raw data of the pilot study. Therefore, the steering angle would be randomly applied between 0° and 2° to keep the vehicle in the middle lane.

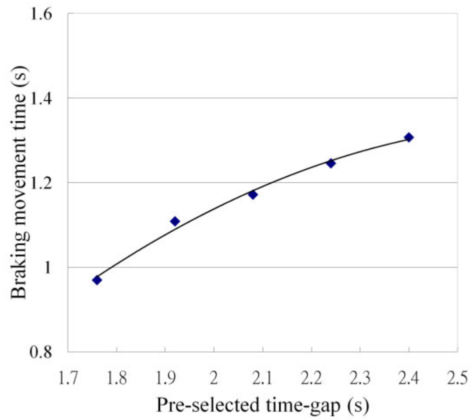


Fig. 3. Function to set up the braking movement time

4 Results

In Fig. 4, the results for both the model simulation and the human drivers (empirical data) were shown. The gray bars meant the results by simulation, and white bars represented empirical data. Also, the comparisons of regressions were shown in Fig. 5 (solid regression lines: simulation results; broken regression lines: empirical results).

4.1 Mean Gap

The model predictions for mean gap showed that as selecting longer time-gaps, the mean gap had obvious increase, no matter driving with or without in-vehicle tasks (Fig. 4). The correlation between simulation and empiric was very high, with the $R > .96$. However, tests indicated that the difference between simulation and empirical results were significant in the no task condition ($F(1,90) = 18.486$, $p < .001$). The difference could be seen in Fig. 5, the linear regression fitting of model prediction and human data. The slope was similar and the difference between regressions was due to the intercept.

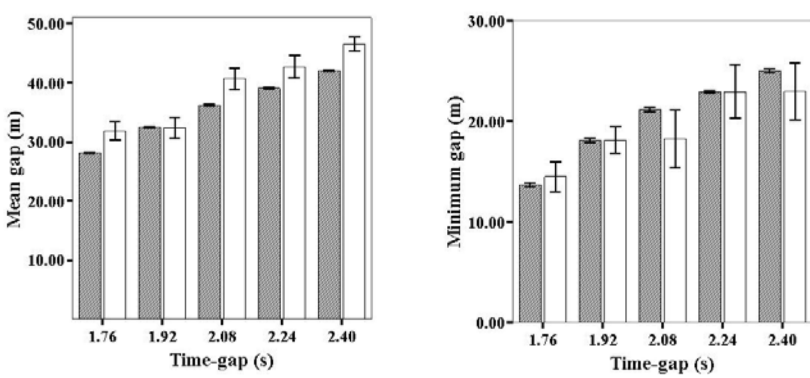


Fig. 4. Means and standard errors from the model simulations and empirical data

4.2 Minimum Gap

The next measure was the minimum gap. Comparing to the human data (as shown in Fig. 4), the overall patterns were similar. The model not only conformed to the basic rank-order effects of variant time-gaps, simulated results were also almost the same to the experimental ones in some conditions. These two types of results were highly correlated ($R > .94$). The difference between them was not statistically significant ($F(1,90) = .606$, $p = .438$). From Fig. 5, two regression lines were close to each other. There was only a little difference existing at the greater time-gap settings.

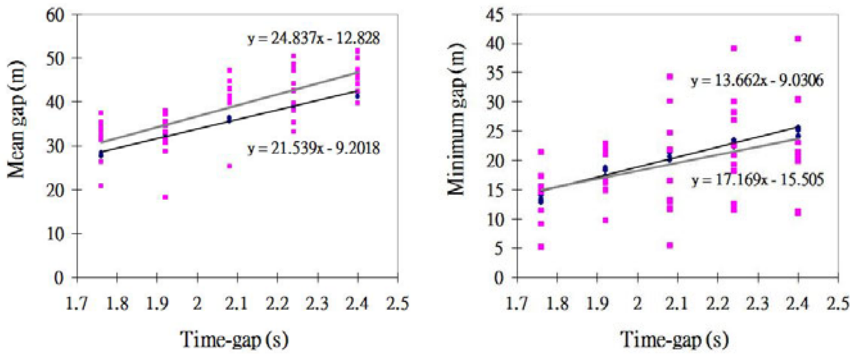


Fig. 5. Linear regressions for model simulation results and empirical data

5 Discussions

In this research, this modeling approach is a conceptual, computational, modular, and simple way to predict effects on driver's performance. The process of modeling in this research follows the following steps: modularization of information processing, building conceptual model with modules and parameter flows, task analysis, simulation, and validation. At the step of modularization, human's behavior is separated into several parts to analyze. For each part, different modules work on their own functions and cooperate as a network. Modules in this study have the same characteristic as what QN-MHP (Feyen, 2002) and ACT-R (Anderson & Lebiere, 1998) stand that modules will work on concurrent activities in parallel.

This human driver model was demonstrated to predict the effects of time-gaps and in-vehicle tasks on driving performance. Several theorem, rules, and observations were referred to construct the model. In the results of model validation, three measures had explained how the model prediction fits the experimental results, respectively. However, a phenomenon occurred that the model works better for driving with in-vehicle tasks, than without task. It may be due to some unnecessary movements when driving. As what ACT-R driver model (Salvucci, 2006) or GOMS (Card, et al., 1983) did, drivers' behavior was assumed to be a fixed sequential movements. But the driver might have some improper actions or unexpected manners, such as trance, daze, impatience, boredom, or even dropping off, especially when they have redundant mental resources. In other words, the positive efficiency from ACC is to save drivers' workload, but it also results in drivers' negligence and omission. These kinds of unimaginable movements and manners are impossible to simulate with modeling approaches. In this research, some subjects indeed have those situations, so that the gap between simulation and empirical data exists.

As to simulating the manual in-vehicle task, the results can accurately predict effects on driver performance. In the experiment, drivers were asked to control the panel for same tasks with identical functions, so their behavior would be more consistent. It is a benefit for modeling because there will be more situations under control. Now it is a simplified model, and validated with ACC evaluation experiments. More measures should be applied to make the validation more complete. Further, some limitations, including internal functions of modules, assumptions, and efficiency of simulation, need to be improved continuously. More general functions in modules must be developed for more cases, such as turning, backing, or overtaking. To improve the efficiency, finite state machine (FSM) and process flow diagram have been applied popularly in simulation (Delorme & Song, 2001; Liu & Özgüner, 2007). These information techniques can systematically represent all the steps, conditions, restricts, and outputs, which enable the process to go smoothly.

6 Conclusions

To create a model for driver behavior simulation, we have searched some other models with different vehicle types, driving scenarios, and modeling techniques. We gather the similarities of these models, including the concept of modularization, queue time, and cognitive architecture, applying to the bus highway driving scenario. This driver model can correctly predict driving performance of mean gap, minimum gap and forward collision rate, especially when drivers need to work on secondary in-vehicle tasks. For auto manufacturers, this modeling approach is very helpful for the phase of ACC prototype evaluation. The interaction of ACC and other vehicle-equipped interfaces, such as navigation devices or some semi-automated assist systems, can be assessed quickly, economically, and accurately. To simulate in other driving scenarios, a task analysis should be performed to develop a new process of behavior for the driving task. Through the way we did, developers can predict and compare the difference among interfaces. Most important of all, by the model simulation, interfaces which are rigorously analyzed will be fewer, so that more efforts can be put into some more complicated field studies.

References

1. Anderson, J.R., Lebiere, C.: The atomic components of thought. Erlbaum, Mahwah (1998)
2. Bellet, T., Tattegrain-Veste, H.: A framework for representing driving knowledge. International Journal of Cognitive Ergonomics 3(1), 37–49 (1999)
3. Card, S.K., Moran, T.P., Newell, A.: The psychology of human-computer interaction. Erlbaum, Hillsdale (1983)

4. Cockburn, A., Gutwin, C.: A predictive model of human performance with scrolling and hierarchical lists. Human-Computer Interaction 24(3), 273–314 (2009)
5. Delorme, D., Song, B.: Human driver model for SmartAHS (No. UCB-ITS-PRR-2001-12). Institute of Transportation Studies, Berkeley (2001)
6. Feyen, R.G.: Modeling human performance using the Queuing Network-Model Human Processor (QN-MHP). Unpublished Ph.D. dissertation, University of Michigan, Ann Arbor, MI (2002)
7. Fitts, P.M.: The information capacity of the human motor system in controlling the amplitude of movement. Journal of Experimental Psychology 47(6), 381–391 (1954)
8. Lin, T.-W., Hwang, S.-L., Green, P.: Effects of time-gap settings of adaptive cruise control (ACC) on driving performance and subjective acceptance in a bus driving simulator. Safety Science 47(5), 620–625 (2009)
9. Lin, T.-W., Hwang, S.-L., Su, J.-M., Chen, W.-H.: The effects of in-vehicle tasks and time-gap selection while reclaiming control from adaptive cruise control (ACC) with bus simulator. Accident Analysis and Prevention 40(3), 1164–1170 (2008)
10. Liu, Y., Özgüner, Ü.: Human driver model and driver decision making for intersection driving. Paper Presented at the 2007 IEEE Intelligent Vehicles Symposium, Istanbul, Turkey (2007)
11. MacKenzie, I.S.: Motor behavior models for human-computer interaction. In: Carroll, J.M. (ed.) HCI Models, Theories, and Frameworks. Morgan Kaufman, San Francisco (2003)
12. Radwin, R.G., Vanderheiden, G.C., Lin, M.-L.: A method for evaluating head-controlled computer input devices using Fitts' Law. Human Factors 32(4), 423–438 (1990)
13. Salvucci, D.D.: Modeling driver behavior in a cognitive architecture. Human Factors 48(2), 362–380 (2006)
14. Salvucci, D.D., Macuga, K.L.: Predicting the effects of cellular-phone dialing on driver performance. Cognitive Systems Research 3(1), 95–102 (2002)
15. Salvucci, D.D., Mandalia, H.M., Kuge, N., Yamamura, T.: Land-change detection using a computational driver model. Human Factors 49(3), 532–542 (2007)
16. Sato, T., Akamatsu, M., Takahashi, A., Takae, Y., Kuge, N., Yamamura, T.: Influence of ACC on frequency of situations experienced while driving on highway. Paper presented at the 16th World Congress on Ergonomics, Maastricht the Netherlands (July 10-14, 2006)
17. Suzuki, H., Nakatsuji, T.: Effect of adaptive cruise control (ACC) on traffic throughput: numerical example on actual freeway corridor. JSAE Review 24(4), 403–410 (2003)
18. Tattegrain-Veste, H., Bellet, T., Pauzié, A., Chapon, A.: Computational driver model in transport engineering: COSMODRIVE. Transportation Research Record 1550, 1–7 (1996)
19. Törnros, J., Nilsson, L., Östlund, J., Kircher, A.: Effects of ACC on driver behaviour, workload and acceptance in relation to minimum time headway. Paper Presented at the 9th World Congress on Intelligent Transportation Systems, Chicago, Illinois (October 14-17, 2002)
20. Tsimhoni, O., Liu, Y.: Modeling steering using the Queueing Network-Model Human Processor (QN-MHP). Paper Presented at the Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting, Denver, CO, USA (2003)
21. Van der Molen, H.H., Bötticher, M.T.: A hierarchical risk model for traffic participants. Ergonomics 31(4), 537–555 (1988)
22. Wu, C., Liu, Y.: Queueing network modeling of driver workload and performance. IEEE Transactions on Intelligent Transportation Systems 8(3), 528–537 (2007)

Beginner Driver Support System for Merging into Left Main Lane

Yuki Nakamura and Yoshio Nakatani

Graduate School of Engineering, Ritsumeikan University
1-1, Noji-Higashi 1, Kusatsu, Shiga 525-0058, Japan
cc012065@ed.ritsumei.ac.jp,
nakatani@is.ritsumei.ac.jp

Abstract. On Japanese high-ways, sharp turn and merging areas are the critical points where the majority of accidents occur. A number of studies have therefore taken place to date on supporting safer driving in merge areas within a framework of Intelligent Transportation Systems (ITS). Most of these studies, however, have assumed that drivers merge into the right main lane from the left lane because it is natural for Japanese drivers to do so, and in actuality those merge areas being rather prevalent in Japan. On actual roads, due to restrictions crossroads or the geographical conditions areas where drivers merge from the right lane into left main lane are also present. Cars in Japan are right wheel drive, thus it is difficult for a driver in a right driver's seat to monitor what is behind on the left. Monitoring left rear involves the motion of turning the body to the left, and at worst, a wheel turn to the left. This may cause a collision with other cars and, even if it is does not result in an accident, a stop or a slowdown, obstructing the following traffic. Beginner drivers find it particularly difficult to merge into the left lane. In this study, we suggest some technical skills that can be of support to beginner drivers when merging into the left main lane. As future work, we discuss a new driver support system that utilizes road-to-vehicle communication equipment via CCD-camera.

Keyword: ITS, merging, main lane, driver support system.

1 Introduction

On Japanese high-ways, sharp turn and merging areas are especially dangerous spots where the majority of accidents occur [1]. Research on support of merging into main lanes has therefore taken place in the past. However, most of that research has involved the expectation that drivers would be merging into right main lane from the left with the reason that vehicles drive on the left in Japan. And in actuality that road structure is quite general. However, actual roads have areas where merging also takes place into the left main lane due to the restriction of crossroads or the situation around them. Right-hand driving makes it difficult for Japanese drivers to check behind them on the left, which is important when merging into main lanes as it can result in other cars having to slowdown or stop and thereby obstructing the smooth flow of traffic. It

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 444–451, 2011.

is also one of the most difficult situations for novice drivers who are not very used to driving [2]. Monitoring left rear involves the motion of turning the body to the left, and at worst, a wheel turn to the left. This may cause a collision with other cars.

In this study we propose a support framework for novice drivers who are not very used to driving cars in safely merging into left main lanes.

2 Related Research

Some studies are related to this theme, including a system of noting the presence of another car passing in order to merge into the main lane via the use of a vehicle detector and DSRC beacons (Fig 1) [3]. Another study provides the degree of danger being displayed on an in-car monitor via the use of a CCD camera attached to the right side of the car and image processing technology [4].

The abovementioned systems are based on software, and both the cost and the time involved can be reduced than using hardware based solutions and can be expected to be put to practical use. However, those systems support cars in the main lane, and do not support cars in a merging lane, many of which involve the premise of cars merging from the left lane into the right main lane.

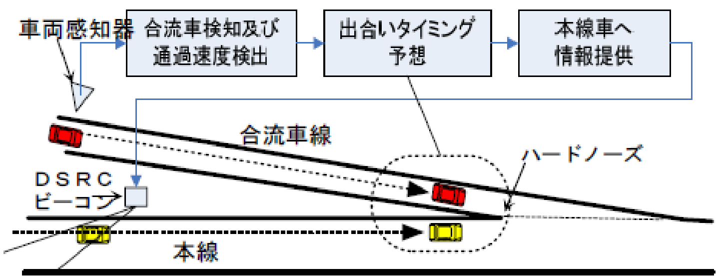


Fig. 1. Basic Framework of Related Studies

3 Outline of Proposed System

(1) External specification of this system. An outline of the equipment setup under consideration is provided in Fig. 2.

With this system a car length and speed sensor is positioned in the main lanes of a high-way and a DSRC beacon in the merging lane. Data measured by the car length and speed sensor is transmitted to the DSRC beacon before being then transmitted to an onboard unit from the DSRC beacon.

(2) Inner specification of the system. This system takes into consideration the data transmitted by the DSRC beacon (the length and speed of other moving cars in the main lane) to an onboard unit that has the data of the length and speed of a merging car itself. The system then suggests whether it is safe to merge or not.

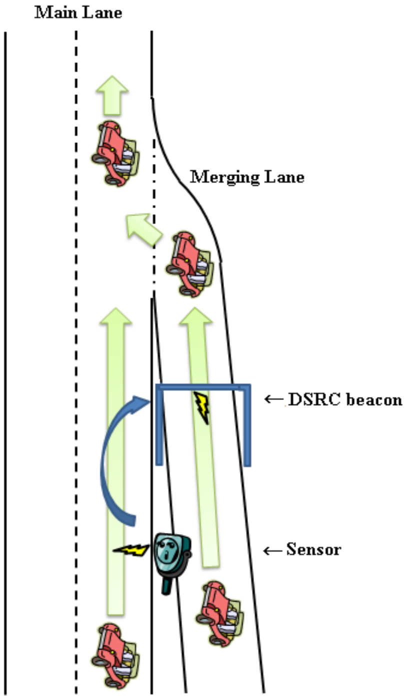


Fig. 2. Outline of equipment used in this study

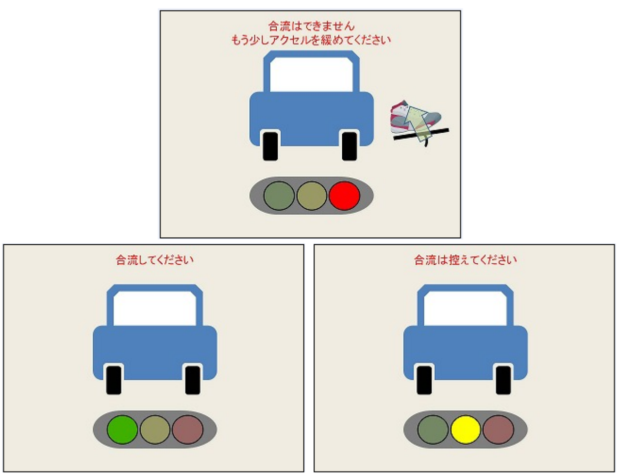


Fig. 3. List of Graphics Presented

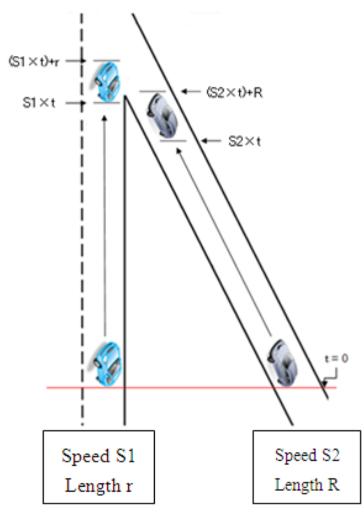
The judgment of merging is shown on the onboard display (Fig. 3). When a merging car approaches a merging area, the onboard unit receives data on the car also approaching to the merging point of the main lane from the DSRC beacon. At this time, the on board system simulates in real-time whether merging is safe using the data from the DSRC beacon and the merging car itself. The system then judges the optimal timing of when to merge and makes the final judgment for the driver. If merging is possible at the current speed, the system displays a green signal. If speed-up or speed-down is required to merge, the system displays a red signal along with a n icon of the gas pedal (speed-up) or the brake pedal (speed-down). During speed-up or speed-down, a yellow signal is on, which urges the driver to pay attention. This support can be expected to make merging easier for beginner drivers.

This judgment is made based on the real-time simulation (Fig. 4).

When a main lane car and a merging lane car are running parallel and the two cars pass the sensor (time t=0), the onboard unit then starts to calculate the travel distance of each car. If a position of the tail of the main lane car compared to the head of the merging lane car reveals no overlap, the system judges merging to be possible.

If there is another car following the main lane car, the system calculates the space between two cars in which the driver can use to merge in safe. Here the main lane car and its follower are supposed to run at the same speed. This is because, if the follower runs faster than the former, the follower will change a lane and pass it, and else if the follower runs slower, the space will widen and merging will become easier.

In addition, the system considers 20m as the safe margin for two cars to merge. This is because it is said that drivers need 0.3 seconds to make a judgment and then to act upon it takes 0.2~0.3 seconds [5], a total of 0.6 seconds. If a car moving at 100km/h, it moves about 17m every 0.6 seconds, and thus 20m is considered to be a safe margin.



$$l_{t+\Delta t} = l_t + S_1 \times \Delta t$$
$$L_{t+\Delta t} = L_t + S_2 \times \Delta t$$

$$l_{t+\Delta t} > L_{t+\Delta t} + R$$

Fig. 4. Information Involved in Real-Time Simulation

Next, suppose that there is a car following the main lane car. In this case, the system reacts differently according to the situation of whether there is a sufficient space for the car to merge or not.

(i) Sufficient space between main lane car and its follower

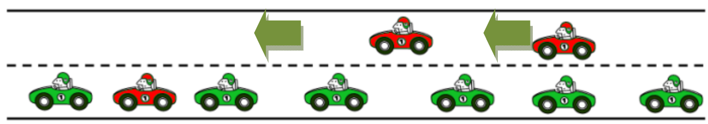


Fig. 5. Merging into enough space between two cars

In this case, the system urges the driver to merge into the space between two cars. The system first displays a red signal to alert the driver and gives a guidance of the speed. When the car adjusts the speed to be in the position where it can merge, the system changes a signal to green.

(ii) Insufficient space between main lane car and its follower

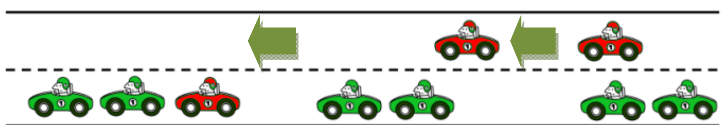


Fig. 6. Merging into insufficient space between two cars

In this case, the system suggests the car slow down so as to merge after the two cars. First the system urges the driver to slow the car down so as to be behind the two cars and displays a red signal before changing it to green.

4 Evaluation Experiment

The system is installed in a car and experimented with it on campus at Ritsumeikan University. The experiments reproduced the virtual merging area (Fig.7). The participants were 10 male students, half of which were novice drivers and the rest drive frequently. The subjective evaluation was gathered by using questionnaires after the drive with the experimental car.

The system was considered effective by all the participants.

Two major opinions were expressed as below.

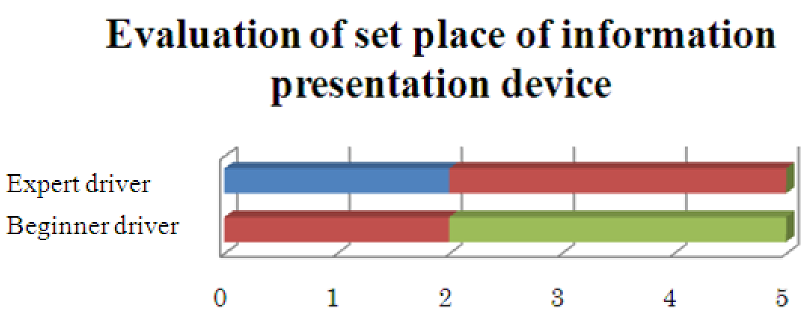
- This system helped alleviate uneasiness when merging
- The graphics are simple and hence we didn't have to scan too much picture content.

Fig. 7. Experimental scene

The defects to be improved were given as below.

- If the system had a sound function it would be more sensory.
- It would be easy for the driver to understand if the tempo of sound changed when approaching to a main lane car.
- Support should be provided not only for merging cars but also main lane cars.

The above opinions were gleaned from the questionnaire.



	Beginner driver	Expert driver
Left	0	2
Center	2	3
Right	3	0

Fig. 8. Evaluation of where the information presentation device should be positioned

A small monitor was used as the information presentation device in the experiment. The participants were questioned on where it would be best positioned so as to be an obstruction while driving. The results are provided in Fig. 8.

The results indicated that novice drivers preferred the right, whereas non-novice drivers preferred the left (in the center).

This is presumably because novice drivers are typically more uneasy with monitoring left rear than non-novice drivers. Novice drivers avoid changing positions, while instead depending on the system and preferring to have the system close to them.

Conversely, non-novice drivers can merge into main lanes without even using the system, and therefore consider it to be one of the aids. They thus think on the left (in the center) to be the best because they can easily check the left side using the rearview.

5 Future Works

We are planning a driving support system that utilizes the road-to-vehicle equipment and CCD-camera. The system uses CCD-cameras instead of DSRC beacons. An outline of the considered set up of the new equipment is provided in Fig. 9. This is because CCD-cameras are more popular in Japanese highways than DSRC beacons.

The current system still has a number of problems. For example, a change in velocity of the main lane cars is yet to be dealt with. CCD-cameras can be used to detect the position and velocity of a car in combination with image processing. This is another reason why CCD-cameras are to be selected in the next step. After performing the image processing the device will transmit the results to onboard units, which will then depict picture content as driving support.

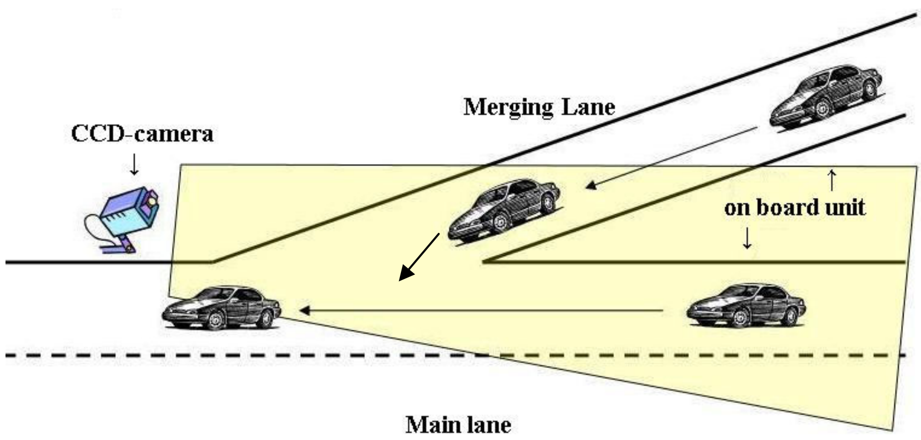


Fig. 9. CCD-camera based traffic detection

In addition, we have to invent an information presentation device, while also reducing the change in center of gravity due to the movement of people's bodies; which is a unique problem with merging into left main lanes from right lanes. With that particular case the driver finds using the rearview to see behind left difficult, and drivers tend to turn the steering wheel using the change in their center of gravity from their body moving. More specifically, we are considering use of a Head-Up Display that makes checking picture content easy when driving a car.

References

[1] Kousoku douro no hashirikata. JAF Mate, 21–25 (2007/8.9) (in Japanese)
[2] Yoshida, Y.: Ease and Safe! Simple Driving Technique, Nihon Bungensya Tokyo (2006) (in Japanese)
[3] Hirai, S., Hatanaka, H., Hirasawa, T., Aya, T., Nishii, S., Nagano, K.: Development of AHS Safe Merging Support Service. In: 6th HIS Symposium, Japan, pp. 331–336 (2007) (in Japanese)
[4] Morita, K., et al.: Merging Support by Image Processing in the Merging Area to Highways. In: 65th Information Processing Society of Japan, pp. 227–228 (2003) (in Japanese)
[5] Matsunaga, K.: Human Science for Traffic Accident Prevention, Nakanishiya Shuppan, Tokyo (2002) (in Japanese)
[6] Japan Safe Driving Center: Research on Night-Time Driving Behavior of Large-Sized Cargo Trucks in the Highways (1999) (in Japanese)
[7] Masakatsu, H., Bagarinao, E., Kurita, T.: Development of Road-to-Vehicle Image Processing Sensor. SEI Technical Review 176, 57–62 (2010) (in Japanese)
[8] Charissis, V., Papanastasiou, S., Vlachos, G.: Interface Development for Early Notification Warning System. In: HCII 2009. LNCS, vol. 5613, pp. 683–692. Springer, Heidelberg (2009)

Multimodal Interface for Driving-Workload Optimization

Hyesun Park[1], Jongwoo Choi[1], Hyeong-Joon Kwon[2], and Kyong-ho Kim[1]

[1] Electronics and Telecommunication Research Institute (ETRI), Daejeon, South Korea
{hspark78,jwchoi,kkh}@etri.re.kr
[2] School of Information and Communication Engineering, Sungkyunkwan University, 300, Chunchun-dong, Jangan-gu, Suwon, Kyungki-do, 440-746, South Korea
katsyuki@skku.edu

Abstract. Today, driving convenience has increased greatly owing to the availability of various telematics devices developed recently. However, this convenience often comes at the cost of driving safety. With the aim of achieving a balance between them, we propose a multi-modal interface for optimizing driving workload and describe an efficient design for the interface. To demonstrate the effectiveness of the interface, we use it in a simulator environment resembling real driving situations, designed to allow the interface to detect and analyze states of both the driver and the vehicle in real time. The proposed interface transfers information by optimizing the driving workload such that it within a range appropriate in view of driver safety. In the future, we intend to demonstrate the feasibility of the proposed interface through multiple experiments.

Keywords: Driving-Workload Optimization, Multi-Modal Interface, Telematics Devices, Vehicle Simulator, Human-Computer Interaction (HCI).

1 Introduction

Today, numerous telematics technologies, i.e., technologies developed by integrating telecommunications with information processing, are applied in vehicles. One such developmental application of this technology to vehicles is to increase the safety or convenience of drivers by providing them with necessary information such as warnings and information on emergencies and traffic situations. However, under certain conditions, there is a high probability of traffic accidents if the driving workload is high [1-2]. Thus, it is necessary to reduce the driving workload through the effective design of interactions between the telematics device and the driver. Among all the information transmitted to a driver driving a vehicle, some information is more important than the other is. Therefore, it is important to prioritize the information transferred to the driver in the case of two or more types of information; then, it would be beneficial to have a tool that would achieve this. One such tool is an adaptive multimodal interface, which facilitates information transfer using an appropriate modality determined on the basis of the driver state. The performance of this tool in prioritizing

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 452–461, 2011.

information and determining modality is affected by various factors such as conditions in the surroundings of the driver, examples of which include movement of vehicles and road situation. In this paper, this movement of vehicles is referred to as the driving task (DT) and the road situation is referred to as the environment task (ET). Using these definitions and concept, we develop a multimodal interface for optimization of the driving workload. The developed interface decides the priority of each type of information transferred to the driver from telematics devices such as mobile phones and navigation systems. Next, it selects one transfer modality from among the visual, auditory, and tactile modalities, and it thereafter classifies and controls that modality's level and transfer time, respectively.

In this study, we validate the effectiveness of the proposed interface by developing an effective design for it. As future work, we intend to verify our results experimentally. As a part of the proposed design, we first create an experimental scenario for each module and build an experimental environment for the scenario. To test the effectiveness of the method that decides the priority of each type of transferred information, the interface first creates a scenario for four driving situations of high importance in real life. Subsequently, we design and construct the interface according to the created scenario. Next, to verify the method that selects modality and its level, the interface transfers the same information to a driver using different modalities. Thereafter, it monitors the driver's feedback and checks her/his driving workload. To this end, the interface simulator is equipped with camera-based eye-tracking equipment and a biological signal measurement device. These devices measure, compare, and analyze the driving workload.

An interface simulator is designed to emulate the actual driving environment. In addition, it sends information to drivers in an actual vehicle at an appropriate time and modality level. Therefore, the proposed interface not only reduces the driving workload but also provides effective information. Furthermore, we expect that the proposed interface will enhance the driver safety. To verify the proposed interface design, we intend to experimentally test it on multiple subjects in the future.

2 Proposed Interface

The proposed multimodal interface consists of four modules: the message input module (MIM), priority decision module (PDM), interaction-managing module (IMM), and message output module (MOM). First, the interface provides more than one type of information as input to the MIM, after which it determines the order of the transferred information using the classified information priority in the PDM. For this, the interface first classifies information with their attribution; subsequently, the information is ordered by weighting importance using a two-task-based classification method. Next, the IMM of the interface selects the most efficient modality for the driver among the visual, auditory, and tactile modalities, and it then controls the modality's level. In this paper, we define the modality level as the transfer strength of modality. Finally, the interface adds any other modality in the absence of feedback from the driver. The proposed interface limits the increase in the driving workload by transferring information using the minimum of a number of modality and its level.

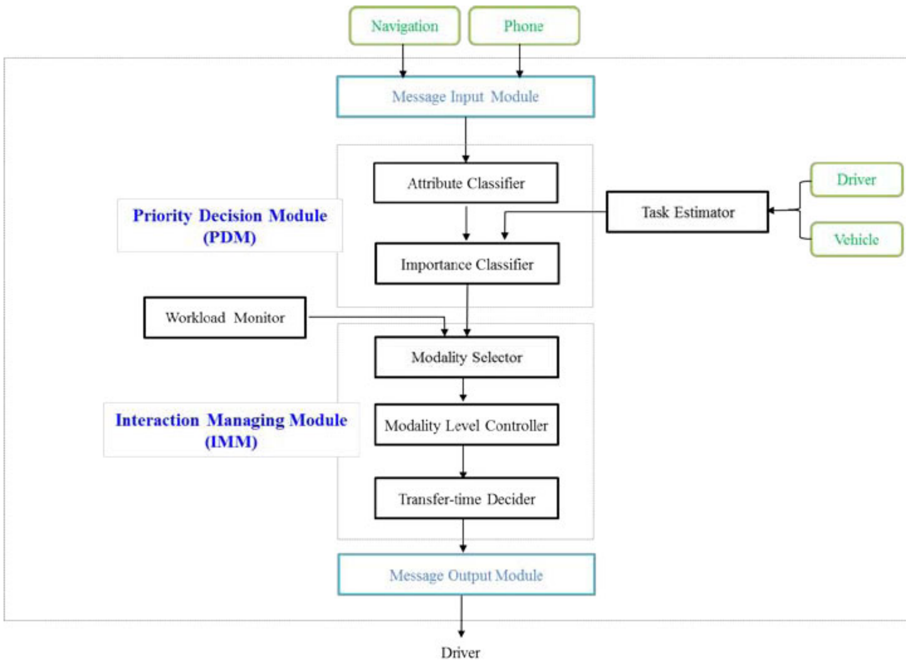


Fig. 1. Proposed interface

The proposed interface, shown in Fig. 1, first provides the PDM with information from a navigation system or phone as input and then orders each type of input information using the corresponding classified attributes and priority weights. Finally, the interface selects an appropriate modality and controls its level with the minimum driving workload.

2.1 Priority Decision Module (PDM)

A vehicle driver may be able to drive more safely and effectively if s/he receives a large amount of information. However, information overload, improper timing of transfer, and unsuitable transferring devices not only increase the driving workload but also interfere with safe driving. Thus, it is essential to first identify the numerous properties of the information given to the driver. Among these properties, the interface classifies the transferred information in terms of its properties of constancy, urgency, and timeliness. Constancy refers to the property that the information is always provided to the driver, whereas urgency is the property that the information provided is of immediate importance to the driver. Timeliness refers to the property that the information should be provided at an appropriate time according to the driving situation. These properties are hereafter referred to as attributes. The proposed interface selects one of these three attributes whenever an input is entered. The interface then classifies both the information transferred from a navigation system or phone and the measured state of the vehicle or driver from sensors into one of the above three attributes (Table 1).

Table 1. Classification results of information transferred from telematics devices to driver and measured state of vehicle or driver into attributes

Class 1	Class 2	Transferred Information	Attribute
Driver State	Safe	Physiological state problem Alarm	Urgency
	Safe	Drowsiness Alarm	Urgency
	Safe	Distraction caused by dialogue in vehicle Alarm	Timeliness
Vehicle State	Safe	CO_2 problem in vehicle Alarm	Urgency
	Safe	Obstacle Alarm	Timeliness
Navigation	Driving	Speed limit Alarm	Timeliness
	Driving	Road rise against over-speed Alarm	Timeliness
	Driving	Rapid turn Alarm	Timeliness
	Driving	Overload vehicle enforcement Alarm	Timeliness
	Driving	Bus zone enforcement Alarm	Timeliness
	Driving	Major accident Alarm	Timeliness
	Driving	Wild animal zone Alarm	Timeliness
	Driving	School zone Alarm	Timeliness
	Driving	Driving Alarm (turn left/go straight…)	Timeliness
	Driving	Driving direction Alarm	Timeliness
	Driving	Entrance Alarm	Timeliness
	Driving	Crossroads Alarm	Timeliness
	Driving	Underground Alarm	Timeliness
	Driving	Highroad Alarm	Timeliness
	Driving	Bifurcation Alarm	Timeliness
	Driving	U-turn Alarm	Timeliness
	Driving	P-turn Alarm	Timeliness
	Driving	Highway Alarm	Timeliness
	Driving	Tunnel entrance Alarm	Timeliness
Phone	Etc.	Receiving a call Alarm	Urgency
		Receiving an SMS Alarm	Timeliness

In the PDM of the proposed interface, the information entered as input is classified into one of the above three attributes as indicated in Table 1. The PDM classifies the obstacles access notification into urgency, whereas it classifies other types of information into timeliness because information entered as input from a navigation system while driving is mostly related to the speed limit or certain special areas such as school zones. Next, in the case of information received through phones, the interface can either activate the SMS service or identify calls that should be made after stopping the vehicle, because the interface prioritizes safe driving. This is explained as follows. If the sender's name is stored in the phone's address book, then the interface identifies that this call is important and subsequently classifies it as an emergency. When the interface determines that the driver cannot take the call immediately, it activates the SMS service and automatically replies to the sender with a message such as "Driving now, will call later."

The proposed interface estimates the ET (environment task) based the current driving situation with driver and vehicle states, which are observed by cameras or other devices, after classifying the information attributes according to Table 1. It can be different to the weighting importance and order of information despite providing the same information. The ETs consist of the vehicle state, driver state, and information transferred from the navigation system and phone; it is affected by surrounding roads, obstacles, and the cost of failing to acquire information about them. DTs (driving tasks) include seven driving actions—left turn, right turn, acceleration, deceleration, U-turn, change from left lane to right lane, and change from right lane to left lane. In reality, actual driving situations are very diverse. Therefore, because it is difficult to estimate each such situation, we select the most typical driving situation that would be encountered while driving and estimate them experimentally in this study. The experimental results are then input to the information classification module. The selected driving situations and the subsequent information provided to the driver are listed in Table 2.

Table 2. Driving situations based on ET and subsequently transferred information

Included Scenarios / Case No.	Driving Situations	Transferred Information
Case 1	Highway, Speeding, Speed camera, Bifurcation, Rear obstacle, Major accident zone	- Bifurcation Alarm - Major accident Alarm - Speed limit Alarm - Obstacle Alarm - Camera speed enforcement Alarm
Case 2	Alley, U-turn, Appearance of a pedestrian, Speed limit region, School zone	- U-turn Alarm - Speed limit Alarm - Driving road Alarm - Obstacle Alarm - School zone Alarm
Case 3	Highway, Drowsiness, Tunnel entrance, Wild animal zone, Lane change	- Drowsiness Alarm - Tunnel entrance Alarm - Lane change Alarm - Wild animal zone Alarm - Obstacle Alarm
Case 4	City road, Hyperventilation, Cold sweats, Work zone, Vehicle blocking, Appearance of a motorcycle, Receiving a call	- Physiological state problem Alarm - Work zone Alarm - Vehicle blocking Alarm - Obstacle Alarm - Receiving a call Alarm

The PDM processes all input information according to the priority determination process shown in Fig. 2. Then, it considers the information attributes and information importance.

As shown in Fig. 2, the PDM first checks whether more than two types of information have been input; if yes, then that information is entered as input, and the priority of each type of information is determined by classifying its attributes and importance. At this point, the database for the weighting importance based on the two tasks (i.e., ET and DT) is used in the method in [3]. Transferred information for each priority is provided to the driver with the most appropriate modality and at a moderate level, in the way described in Section 2.2.

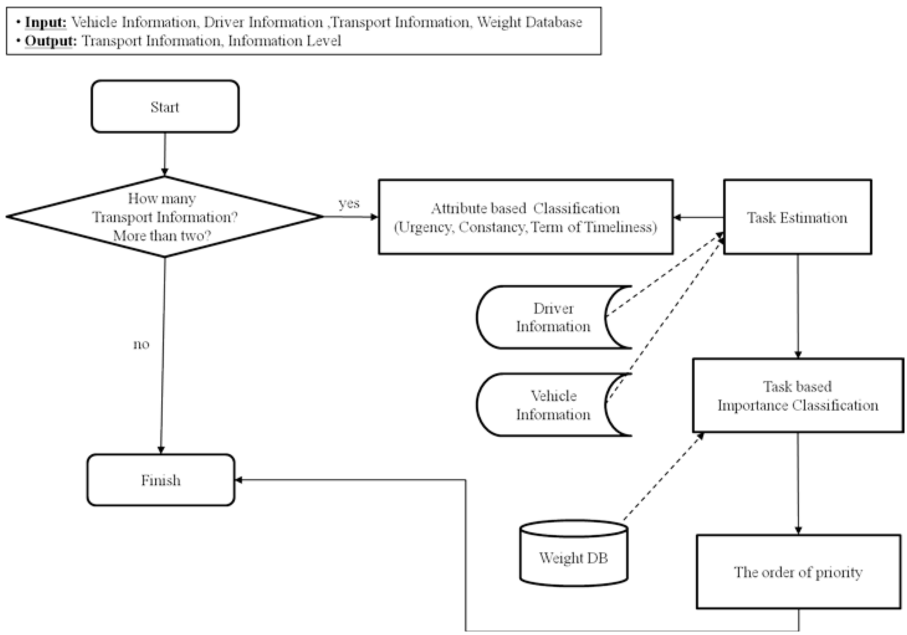


Fig. 2. I/O processing sequence in priority decision module (PDM)

2.2 Interaction-Managing Module (IMM)

This module not only reduces the driving workload but also effectively provides transferred information by determining the means of transferring "what, how, and when," after the priority of the information entered as input is determined. The proposed interface transfers information to the driver using one out of the following three modalities: visual, auditory, and tactile. By default, navigation information is provided through the visual and auditory modalities because these modalities are most efficient. However, in certain circumstances, these two modalities divide. Drivers get nearly 90% of their information about the driving situation through the visual modality [4]. Nevertheless, when driving, drivers still get the feeling of a high driving workload because they must maintain a frontward gaze. Consequently, it is very important to select the appropriate modality to ensure efficient transfer and driving safety. For this purpose, the IMM first selects a modality by comparing the visual and auditory modalities. Subsequently, it supplements the selected modality with additional modalities or changes the level of the modality by checking the driver's cognitive ability. To this end, our interface efficiently provides information such that drivers get the feeling of minimal driving workload. Fig. 3 describes the I/O processing sequence in the IMM.

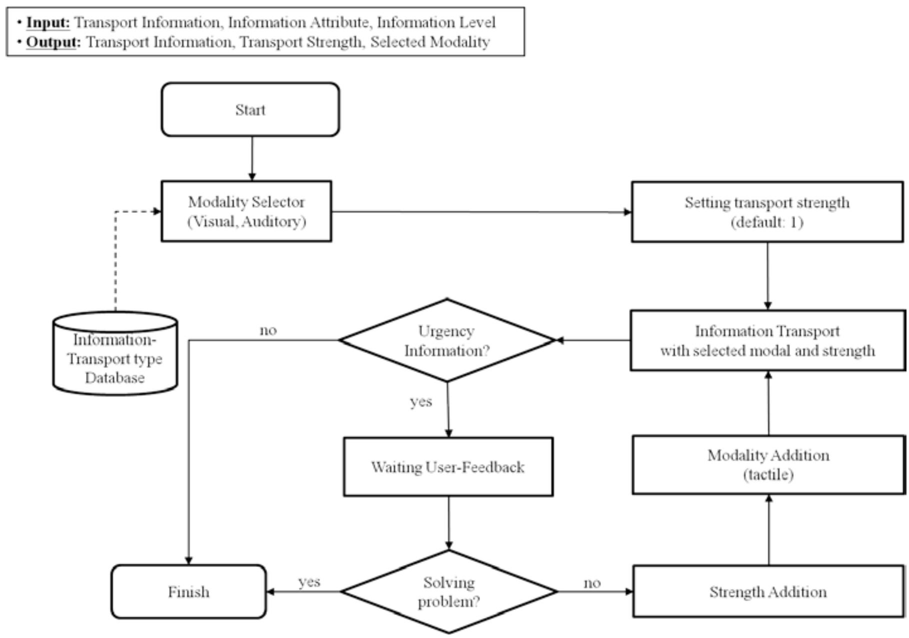


Fig. 3. I/O processing sequence in interaction managing module (IMM)

First, the IMM selects the modality on the basis of each measured driving workload from among the visual, auditory, tactile, and cognitive modalities when the information is input. As stated previously, the IMM compares the visual and auditory modalities and then selects the modality with a lower workload. The selected modality is kept at the minimum level by default, after which the IMM changes the level according to the importance of the transfer and the results of the observed driver feedback. Thus, the level of the selected modality may gradually increase; alternatively, other modalities may be added. To achieve this, the IMM detects the states of the driver and vehicle in real time and monitors the driver's feedback with changes in these two states. For example, when the system detects drowsiness, it sounds an alarm until the driver recognizes it and wakes up. At this point, the system detects the driver's state using a camera-based device via three measures related to the driver's eyes. The first measure is variations in pupil size (i.e., pupil opening); this measurement is an effective approach for detecting the workload [5,6]. The second measure is eyelid movement, whose measurement is useful in determining driver fatigue [7,8]. For this measurement, the IMM uses a measure very similar to the percentage of eye closure (PERCLOS). To calculate this measure, the IMM determines the period for which the driver does not blink. In this study, we determine the driver's drowsiness using PERCLOS. The third measure is the gaze trail (trajectory), which is used to determine the extent of the driver's distraction. In our simulator, we use faceLab-5 as the gaze estimation device. To detect the driver's eye, the threshold of 0.75 is set for the device, based on experimental results. Generally, if the driver's gaze departs from a predetermined area, then the IMM determines that s/he is distracted while driving.

On the other hand, we also detect and measure the state of the vehicle using the methods established in [9].

3 Experimental Environment and Design

To verify the effectiveness of the proposed interface, a real simulator environment as shown in Fig. 4 is set up. The left side of the figure shows the actual simulator environment. A detailed explanation of this experimental environment is given below.

Fig. 4. Experimental environment of the proposed interface

3.1 Driving Simulator

The simulator was developed in a laboratory at Kookmin University[1]. The proposed interface was operated in this simulator environment. The simulator has a pc-based full-scale environment with three front channels and one rear channel. Therefore, it can generate realistic and highly graphic images and simulate vehicle movements resembling those in real driving with roll and pitch. In addition, it is equipped with a monitoring system, a device for measuring physiological signals, and an eye-tracking device. We describe them in detail below.

a) Physiological signal measurement device

For measuring the physiological signals, we use the MP100 data acquisition system (BIOPAC, USA), which can detect and analyze various physiological signals using up to 16 channels. The system consists of MP100A-CE and UIM 100c, which are the data acquisition unit and universal interface module, respectively. The system has various cables and USB adapters. We use Acqknowledge-3.7 as the analysis program.

---

[1] This work was partly supported by the Vehicle Control Laboratory in the Graduate School of Automotive Engineering at Kookmin University in Korea.

b) Eye tracking device

For eye tracking, we use the faceLab-5 system developed by Seeing Machines, Australia. The system extracts gaze direction and estimates eye and head movements. In addition, it detects the pupil size, eye opening, and PERCLOS. The system consists of two digital cameras and one infrared illumination. We operated an additional camera as the scene camera. Data were acquired and analyzed through data synchronization with connections to two note-PCs using IEEE 1394 and USB interfaces.

4 Conclusions and Future Directions

Owing to the availability of a multitude of electronic automotive devices developed in recent times, it is now possible to obtain more information for ensuring convenience and decreasing monotony for drivers. However, these opportunities may at times pose a threat to driver safety, because increase in information received causes an increase in the driving workload. The proposed multimodal interface optimizes the driving workload and controls the information transferred to the driver by considering the driver's state using various measurement devices. The proposed interface is designed with a real driving simulator, and it analyzes and detects the states of the driver and vehicle using the setting devices. Subsequently, the interface estimates the workload and the driving tasks. In addition, it determines the modality, level, and timing of transferred information using the estimated workload. In a future study, we intend to experimentally test the effectiveness of the interface in various other driving situations and involving a large number of drivers. Then, if the interface proves to be effective, we can aim to equip cars with it so as to make them driver-oriented futuristic intelligent cars.

Acknowledgments. This work was supported by the Industrial Strategic Technology Development Program and the Development of Driver Adaptive Intelligent HVI (Human-Vehicle Interface) Technology (10033346) funded by the Ministry of Knowledge Economy (MKE, Korea).

References

1. Koo, T.-Y., Park, K.-J., Kim, B.-Y., Kim, H.-J., Suh, M.-W.: A Study on Driver's Workload of Telematics using Driving Simulator - A Comparison among Information Modalities. International Journal of Precision Engineering and Manufacturing 10(3), 59–63 (2009)
2. Road Traffic Safety Authority (South Korea): Statistical Analysis of Traffic Accident: RTSA Annual Report, pp. 61–70 (2005)
3. Kwon, H.-J., Kim, D.-J., Park, H.-S., Kim, K.-H., Hong, K.-S.: A Method for Evaluation of in-vehicle Information Level for Driving Workload Optimization. In: Proc. of the Summer Conference of Korea Institute of Signal Processing and Systems, vol. 11(1), pp. 178–181 (2010)
4. Cohen, A.S.: Feed Forward Programming of Car Driver's Eye Movement Behavior - A System Theoretical Approach: Final report volume one (1980)

5. Tyeruna, L., Gleckler, M., Stoltzfus, D., Johnson, S., Goodman, M., Wierville, W.: A Preliminary Assessment of Algorithms for Drowsy and Inattentive Driver Detection on the Road: Tech. Rep. DOT HS 808 (1999)
6. Ahlstrom, C., Kircher, K., Kircher, A.: Considerations when Calculating Percent Road Centre from Eye Movement Data in Driver Distraction Monitoring. In: 5th International Driving Symposium on Human Factors in Driver Assessment, Big Sky, MT, US, pp. 132–139 (2009)
7. Palinko, O., Kun, A.L., Shyrokov, A., Heeman, P.: Estimating Cognitive Load using Remote Eye Tracking in a Driving Simulator. In: Proc. of the 2010 Symposium on Eye-Tracking Research & Applications, ETRA 2010. ACM, New York (2010)
8. Hanowski, R.J., Bowman, D., Alden, A., Wierwille, W.W., Carroll, R.: PERCLOS+: Moving Beyond Single-metric Drowsiness Monitors: SAE Technical Paper Series, Commercial Vehicle Engineering Congress & Exhibition (2008)
9. Yoon, D., Choi, J., Kim, K., Kim, H., Narayanan, H.: Attention Based Driving Event Simulator for Telematics Environment. In: Proc. of the International Conference on Advanced Communication Technology, vol. 1, pp. 757–759 (2009)

Part V

Social and Environmental Issues in HCI

Proposal of a Method for Promotion of Continuous Pro-Environmental Behavior with Easy Communication

Saizo Aoyagi, Tomoaki Okamura, Hirotake Ishii, and Hiroshi Shimoda

Graduate School of Energy Science, Kyoto University, Yoshida-honnmachi, Sakyo-ku, Kyoto 606-8501, Japan
{aoyagi,okamura,hirotake,shimoda}@ei.energy.kyoto-u.ac.jp

Abstract. In Japan, Promotion of Domestic Pro-Environmental Behavior(PEB) is one of the main challenges for energy saving. Existing studies of Promotion of PEB seldom deal with continue of PEB. Purposes of this study are proposal of a method for promotion of continuous PEB with easy communication and evaluation of the effectiveness of the method. Main part of the proposed method is an easy communication system to cause social facilitation of PEB among users. An evaluation experiment with ten participants was conducted and the result showed that the proposed method promotes domestic PEB if "feeling of joint PEB" is aroused.

Keywords: Pro-Environmental Behavior, Computer Mediated Communication, Social impact theory, Easy communication, Social stress.

1 Introduction

In Japan, energy consumption of residential sector has risen because both number of households and households energy consumption have increased. As a result, energy saving in this sector is recognized as one of pressing issues by the government [1]. Domestic Pro-Environmental Behavior (PEB) of people is a concrete way of energy saving in residential sector and there have been many studies about factors which promote PEB of people [2, 3] and methods of promotion of that [4-6]. Most studies in this area have just deal promotion of PEB, and few studies [7] definitely focused on continue of PEB despite of its importance. Promotion of starting of PEB of people has no meaning if they give up PEB easily.

Purposes of this study are proposal of a method for promotion of continuous PEB with easy communication and evaluation of the effectiveness of the method.

2 Method for Promotion of Continuous Pro-Environmental Behavior

2.1 Approach and Requirements of the Proposed Method

A study said that habit formation of an action needs continuance of that action about a month [9]. Based on the study, we intend to continue PEB of target people of the

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 465–473, 2011.

method until habit formation of PEB, and the method is designed to be used for a month, not endlessly used.

Next, this study employed socio-psychological social facilitation [8] effect as a main principle for promoting continuous PEB. Social facilitation means that people conform their action to others. This effect works on people each other, and will form feedback loops of PEB; social facilitation therefore will be useful to promote continuous PEB.

- Stress-free communication about PEB

Nevertheless, some study said that people recently do not like conventional deep or close communication, because of social stress associated with it and stress-free is main point of continue communication [10, 11]. Recent development of twitter, very easy way of communication supports these studies [12]. This is a problem because knowing the fact that others do PEB is essential to cause social facilitation and it is conventionally associated with communication. Consequently, stress-free and easy communication method for PEB promotion is necessary.

- Feeling of joint PEB in the same "room"

Nevertheless, it is afraid that easy and stress-free communication causes weaker social facilitation because social impact can be a kind of social stress. Therefore, social facilitation caused by easy communication is desired to be as storing as other requirements allow.

- Secure existence of others doing PEB

Next, it is possible that the feedback loop of social facilitation of PEB is broken when a user feel that others don't do PEB by some chance. Consequently, a measure for avoiding break of the feedback loop of social facilitation of PEB is needed.

- Awareness of PEB occasion

At last, according to cognitive psychology, people's actions in daily life are embedded in time-series schema and implemented following schemas in each situation without being intended [14]. Therefore, if someone does not do PEB in daily life, occasions of PEB in daily life do not surface to the conscious mind and are likely let past. Consequently, some way to show occasions of PEB to people who do not do PEB in daily life in order to promote PEB of them.

2.2 Details of the Proposed Method

We propose a method for promotion of continuous PEB. The proposed method employs an easy communication system which uses portable multimedia device or Smartphone, its client software, web server and server software. Users of the system communicate in a group and are promoted their continuous PEB. The system has following functions. Screenshots of the proposed system are shown in fig 1.

- Easy communication method

Main functions of the system are two easy communication methods, "mutter" and "PEB footprints" which clear some kinds of social stress for sending information about PEB which was done.

"Mutter" is a limited text-based Computer Mediated Communication (CMC) method shown in right part of fig. 1. It has no destination address; it therefore cannot be use for one-on-one communication in order to free users from obligation of reply and expectation of replies, which can be a social-stress. In addition, the number of characters of mutter is limited up to 140 in order to free users from obligation of writing "good text" with caring about others' eyes.

"PEB footprint" is another CMC method shown in left part of fig. 1.

It can be sent through one-touch operation, and expresses only the fact that someone does a PEB by fixed sentence for each PEB. Therefore users need not write text by themselves in order to free users from obligation of writing "good text" with caring about others' eyes.

- Room-specific timeline

Next, "room-specific timeline" displays others' mutters and PEB footprints to a user shown in left part of fig. 1. According to social impact theory [13], immediacy affects impact of social facilitation. Based on this theory, "room-specific timeline" promotes social facilitation effects through strengthening spatial immediacy artificially and arouse a kind of feeling of users. They will feel that they do PEB together in the same "room" when they view room-specific timeline of the place they are in that time. This effect is named as "feeling of joint PEB in the same room". This feeling do not mean that people think physically occupy the same place. "Room" in this context is just an imaginings.

- Conversational agent

In addition, "Conversational agent" promotes formation of positive feedback loop of social facilitation through behaving as "others who do PEB". Conversational agents always behave as artificial active people who do PEB in order to avoid such situations. It is implemented with artificial intelligence technologies or acted by people who are in charge of promotion of PEB of people.

Fig. 1. Screenshots of the proposed method

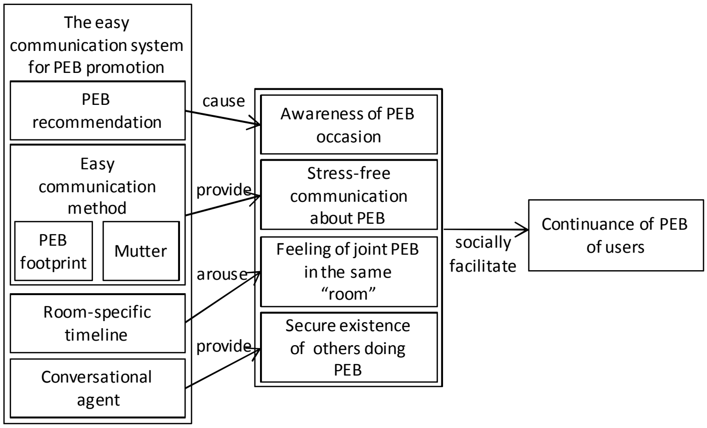


Fig. 2. A framework of the proposed system

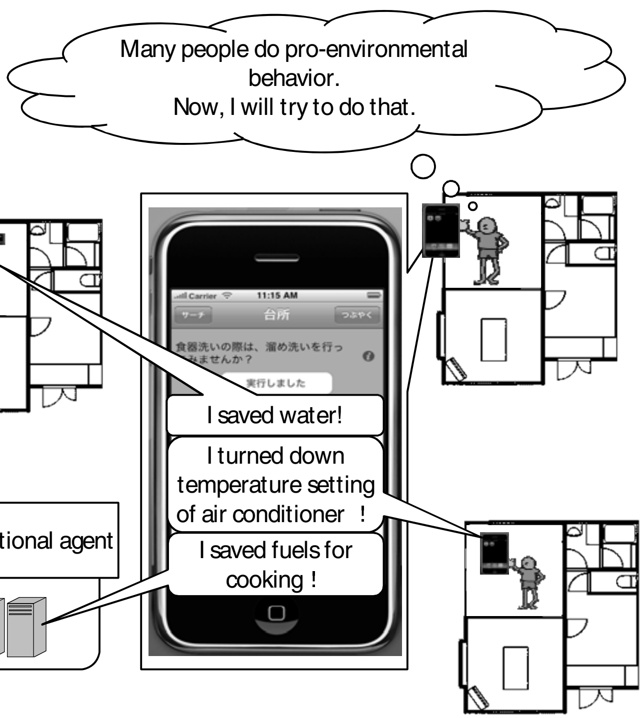


Fig. 3. An example of how the proposed system works

- PEB recommendation

At last "PEB recommendation" gives awareness of occasions of PEB of people based on time schedule of their daily life so as avoid oversight of the occasions. It shows an available PEB in that time to users as shown in left part of fig. 1.

Correspondences between requirements and functions are shown in fig. 2.

Figure 3 shows an example of how the proposed system works and how users use it. As this figure shows, users carry around smartphone with the client software of the system in their home, and announce the fact that they do PEB to other users with “PEB footprints” function of the system. “PEB footprints” are shown to other users by “room specific timeline” functions of the system in displays of smartphone. “PEB footprints” are sometimes sent by conversational agents in order to avoid break of the feedback loop of social facilitation. If some users view many “PEB footprints” in “room specific timeline”, they will intend to do PEB by social facilitation effect.

3 Evaluation of the Proposed System

3.1 Purpose

An evaluation experiment was conducted for evaluation of effectiveness of the proposed system through checking following two points. First is whether feeling of joint PEB is aroused and second is promotion of continuous PEB by feeling of joint PEB.

3.2 Method

In the experiment, some participants actually used the system in their households, and the period of use in the experiment was a month from 2010 November 15 to December 14. Participants of the experiment were two groups. The younger group consists of five 20's people and the elder group consists of four 50’s people and a 60 person. In addition, each group includes five conversational agents. Each participant used an iPod touch which client software of the system was installed in. Each iPod touch was connected to server software of the system through Wi-Fi station.

Before the period of use, a questionnaire was conducted to gather information about participants’ daily life pattern, frequency of PEB, and other basic information of them. In the period of use, participants use the client software in iPod touch and communicate with other group members, 58 kinds of PEB were recommended based on their daily life patterns. In addition, they answer questionnaires about frequency of PEB for the last week, which were recommended by the system once a week. After the period, a questionnaire was conducted to gather participants‘ subjective evaluation of starting, increasing frequency, and continuing of PEB through use of the system, and subjective evaluation of functions of the system, such as whether they felt others doing PEB in the same "room" through use of the system. Participants got ten thousand Japanese yen after all activity in the experiment were finished. They were told about this reward before the period, they however were not told amount of money.

Log data of the system, such as number of PEB footprints of each participant were recorded in the period. This log data and answers of questionnaire were collected in order to evaluation of the system from multiple points of view.

3.3 Results and Discussion

All participants answered all questionnaires and log data was collected without problems. Table 1 shows results of questionnaire on effectiveness of system functions. Y1-Y5 mean participants in the younger group and E1-E5 mean participants in the elder group. This table shows that five participants felt (answered 4 or 5) others doing PEB in the same "room" and remaining participants felt others doing PEB not in the same “room”. Moreover, most participants didn’t feel (answered 1 or 2) four social-stress associated with communication, “minded others”, “expected replies”, “felt obligated to reply”, “felt obligated to reply to reply”. These results suggest effectiveness of functions of the easy communication system.

Table 1. Results of questionnaire on effectiveness of system functions

Participant	Felt others doing PEB in the same "room"	Felt others doing PEB**	Minded others	Expected replies	Felt obligated to reply	Felt obligated to reply to reply
Y1	5		1	2	4	2
Y2	2	5	1	3	2	3
Y3	2	4	1	1	1	1
Y4	4		2	1	1	1
Y5	2	5	2	4	4	3
E1	2	3	4	1	1	1
E2	5		1	3	4	2
E3	1	4	1	1	3	1
E4	4		2	4	2	2
E5	4		2	2	2	3

* Answered with number (5: I completely agree - 1: I completely disagree)
**: Participant who did not felt others doing PEB in the same "room" only.

Figure 4 shows frequency of PEB of participants. Participants answered “always”, “often”, “sometimes”, “occasionally”, “not at all” and “no occasion” about 58 PEBs to questionnaires which measures frequency of PEB. PEBs which were not answered as “no occasion” through the period by each participant were shown in fig. 4 for comparison. Figure 4 shows that frequencies of PEBs had increased basically.

Nevertheless, there were some exceptions. For example, PEB frequencies of Y1, Y3, Y4, E1, E4, and E5 once decreased from the questionnaire before the period to the questionnaire a week after, and increased after that. These results are explained by difference between questionnaire before the period and other questionnaire of PEB frequency. The questionnaire before the period is considered to be answered by guess because people were assumedly not conscious of their behaviors in daily life, other questionnaires however are considered to be answered based on credible memories because people were conscious of PEB being affected this experiment. Consequently we mainly use results of questionnaires from a week after to a month after, and results of the questionnaire before the period are used as reference.

Next, Kruskal-Wallis tests were conducted in order to evaluate promotion of PEB through the period and table 2 shows significance of change of PEB frequency through the period. Considering the phenomenon described above, Kruskal-Wallis tests were conducted (1) between answer of the questionnaire before the period and that of a month after, and (2) between that of a week after and that of a month after. Table 2 shows that five participants had increased PEB frequency significantly through the period and three participants in them had evidently increased PEB frequency.

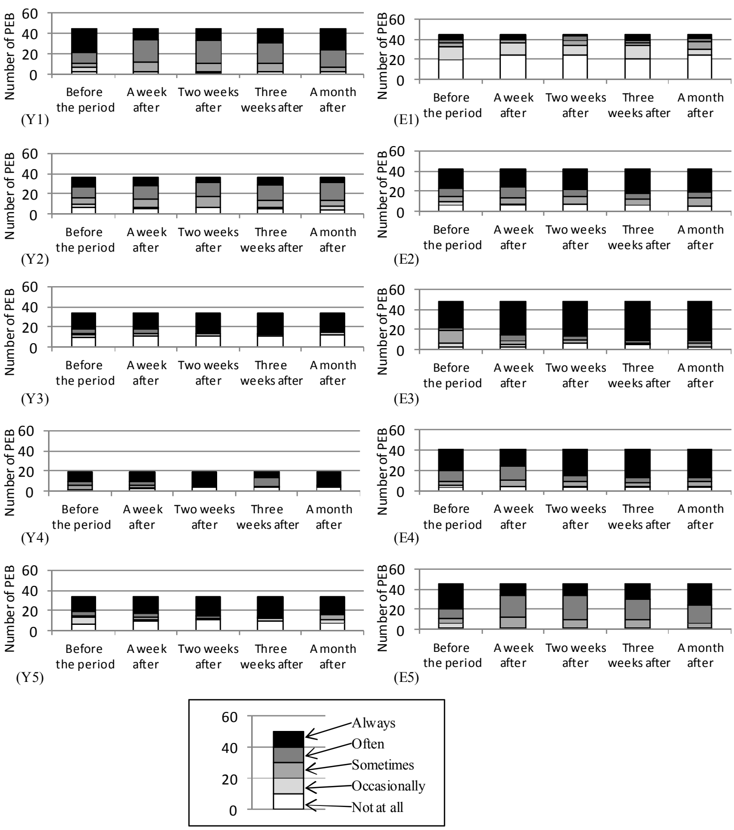


Fig. 4. Results of questionnaire on frequency of PEB

Figure 5 shows the number of "PEB footprints" in the experimental period. PEB footprints of participants showed decreasing trend, they however continued through the experimental period. It is evidence which suggests effectiveness of the system.

Next, table 3 shows subjective evaluation of starting, increasing frequency, and continuing of PEB through use of the system. This table shows that four participants began new PEB, six people increased PEB frequency, and five participants continue PEB subjectively.

To summarize, fig. 2, table 2 and table 3 suggest that PEB of about a half of all participants were promoted and continued a month.

There are some differences between results of questionnaire on PEB frequency shown in fig. 2 and subjective evaluation after a month and shown in table 3. For example, E3 could not begin new PEB, increased PEB frequency, and continued PEB subjectively, he/she however show significant increase of PEB frequency. Neverthe-less, they are considered to be caused by lapse of memory or change of baseline of evaluation, which are common thing to human.

At last, we found a factor which affected promotion of PEB and continuous PEB in comparison between table 1 and table 3.Table 1 showed that Y1, Y4, E2, E4, and E5 felt (answered 4 or 5) others doing PEB in the same "room" and the same participants

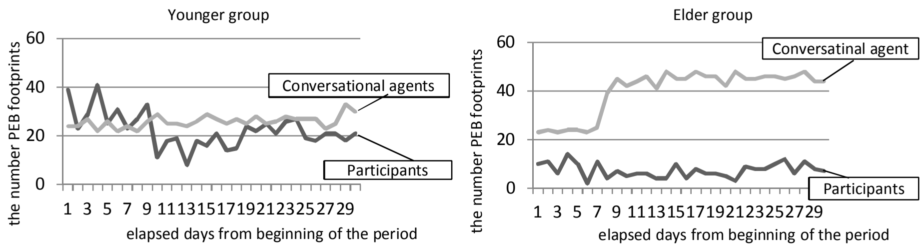


Fig. 5. The number of "PEB footprints" in the experimental period

Table 2. Significance of change of PEB frequancy between two questionnaires

Participants	Before the period and a month after	A week after and A month after
Y1	n.s.	*
Y2	n.s.	n.s.
Y3	n.s.	n.s.
Y4	n.s.	n.s.
Y5	n.s.	n.s.
E1	n.s.	n.s.
E2	*	n.s.
E3	**	n.s.
E4	n.s.	**
E5	n.s.	*

*: p<0.05, **:p<0.01.

Table 3. Subjective evaluation of starting, increasing frequency, and continuing of PEB through use of the system

Participant	Began new PEB	Increased PEB frequency	Continued PEB
Y1	3	5	5
Y2	3	4	2
Y3	1	3	2
Y4	4	5	4
Y5	1	2	2
E1	2	2	2
E2	5	4	4
E3	2	2	2
E4	4	4	4
E5	4	4	4

* answered with number (5: I completely agree - 1: I completely disagree).

also increased continued PEB (answered 4 or 5) in table 3. In addition, about the same participants increased PEB and begun PEB (answered 4 or 5). These results suggest feeling of joint PEB in the same "room" is a key factor to promote continuous PEB and only feeling of joint PEB (not in the same "room") is not sufficient.

4 Summary and Future Task

In this study, the method for promotion of continuous PEB with easy communication was proposed and it was evaluated through the experiment. The results showed that continuous PEB of some participants were promoted by the proposed method.

Nevertheless, other participants however not continued, and suggested that feeling of joint PEB in the same "room" is a key factor to promote continuous PEB. In this regard, however feeling of joint PEB was achieved in remaining participants. This result suggests that remaining assignments for us is to enhance immediacy which is a factor of strengthening social facilitation effect. In other word, future task is to enhance feeling of being in the same "room".

References

1. METI: Annual Energy Report 2010 (2010)
2. Hirose, Y.: Determinants of Environment-Conscious Behavior. Social Psychology 10(1), 44–55 (1994)
3. Kollmus, A., Agyeman, J.: Mind the Gap: why do people act environmentally and what are the barriers to pro-environmental behavior? Environmental Education Research 8(3), 239–260 (2002)
4. Abrahamse, W., Steg, L., Vlek, C., Rothengatter, T.: A review of Intervention Studies Aimed to Household Energy Conservation. Journal of Environmental Psychology 25, 273–291 (2005)
5. Ohta, K., Fujii, S.: An Experimental Study on Effects of Providing Objective Information about Reduction of CO_2 Emission due to Pro-Environmental Behavior. Journals of the Japan Society of Civil Engineers G. 63(2), 159–168 (2007)
6. Matsumura, N.: A study on Effects of Role Playing on Behavioral Change toward Environmental Conservation. In: Proceedings of Infrastructure Planning, vol. 27, p. 144 (2003)
7. Tan, Y.: Persuasive Technology in Motivating Household Energy Conservation. In: Michaelles, F. (ed.) Business Aspects of the Internet of Things Seminar of Advanced Topics. FS 2009, pp. 52–58 (2009)
8. Moscovici, S., Personnaz, B.: Studies in Social Influence. Journal of Experimental Social Psychology 16, 270–282 (1980)
9. Sharma, R.: Who Will Cry When You die? HarperCollins, Canada (2000)
10. Sato, I.: Communications of "shyness" in dance: body presentation of "I" in club. Journal of Communication Studies 27, 51–72 (2007)
11. Muramoto, Y.: Fuzzy Boundary between Group and Togetherness Situation: A Variety of Identity Found at a Park in Every Morning. Japanese Journal of Social 12(2), 112–124 (1996)
12. Twitter, `http://www.twitter.com`
13. Latane, B.: The Psychology of Social Impact. American Psychologist 3(4), 343–356 (1981)
14. Schank, R.C., Abelson, R.P.: Scripts, Plans, Goals, and Understanding: An Inquiry into Human Knowledge Structures. The Halsted Press Division of John Wiley and Sons, New York (1977)

A Context Centric Approach to Utilize Social Media Services on Public Terminals

Micha Block[1], Jasmin Link[1], and Simon Thiel[2]

[1] University of Stuttgart - Institute for Human Factors and Technology Management (IAT), Nobelstr. 12, 70569 Stuttgart, Germany
[2] Fraunhofer-Institute for Industrial Engineering (IAO), Nobelstr. 12, 70569 Stuttgart, Germany
{micha.block,jasmin.link}@iat.uni-stuttgart.de, simon.thiel@iao.fraunhofer.de

Abstract. The development of new technologies for touch sensitive surfaces led to a revival of touch interfaces in innovative scenarios and a wide field of application. They can now be found in a range of hardware products on the market from small handhelds and tablets, to terminals, interactive tables and walls. In combination with information and communication technologies like GPS, wireless LAN, cameras, motion and light sensors – as most of the smart phones are already equipped with – the way is set for modern social software to conquer these devices. The paper illustrates a context centric approach to utilize these new technologies to access social media services in public location-based contexts on "Social Terminals".

Keywords: public social terminals, touch interfaces, multitouch, context centric design, social media services, terminal development, concept evaluation.

1 Introduction

Social media became a vital part of both leisure and business world. Accessible through web platforms – location independent – their potentials grow for utilization in location-based, public contexts on specific hardware devices. Public terminals currently experience a revival because of their new display and touch technologies and innovative construction concepts. These devices allow an access of social media services location-based in a rich user experience. Moreover, they facilitate to embed social media services and their related business models into collaborative scenarios, in which users work, plan and innovate together.

This paper describes the application of a context centric approach for the development and evaluation of "Public Social Terminals" that aggregate the benefits of both social media concepts and the potentials of public terminals, as a location-based access for their user group. It describes a step-by-step process how to utilize the key benefits and best practices of a context with its specific business models, user groups and application services, and adapt it to a specific target context. The approach has a strong focus on reusing existing concepts of the source context by adapting it to a specific target context. In this case the source context is the Social Semantic Web

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 474–482, 2011.

with its valuable services for users, its networking, communication and information capabilities, and its rich media content and user experience. To exemplify the approach, as a target context the scenario of a “public conference” has been chosen where stationary terminals aggregate social media services to support the participants and conference management.

The approach was developed based on existing methods for and expertise in the development and evaluation of both social media platforms and touch-interfaces by the Fraunhofer Institute for Industrial Engineering (IAO) in Germany, in cooperation with the Institute for Human Factors and Technology Management (IAT) of the University of Stuttgart. The approach can mainly be used in two phases during the development process: Firstly, the early conceptual design and secondly, for the evaluation of existing interactive systems. In contrast to Human Interface Guidelines (HIGs) for touch devices and frameworks – like Apple´s HIG for their mobile touch devices [1] or Microsoft´s Surface development framework [2], this approach does not focus on graphical design and general interaction guidelines. It presents an approach how to adapt business models and best practices of applications of a source context to a specific target context. Following this, it adds a development process for the resulting concept using device specific design and interaction patterns, as well as sets of hardware specifications.

2 The Approach in Five Dimensions

The process of adopting and adapting from one to another context comprises five dimensions: the source and target context, the business models, the specific user groups, the application features and services (best practices) and the requirements to hardware devices. The following sections describe these dimensions in detail:

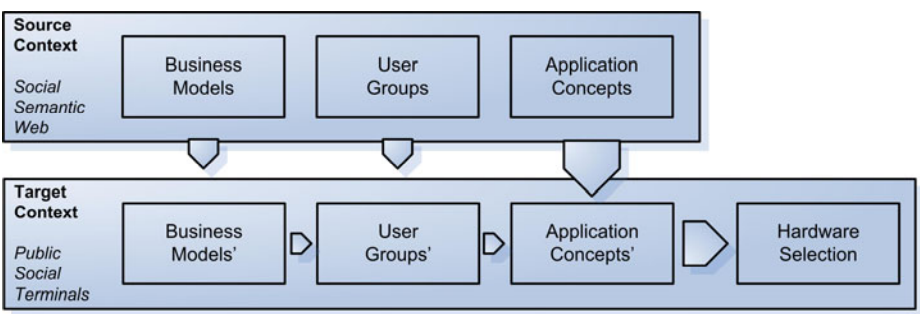


Fig. 1. The contexts comprise business models, user groups and applications concepts which serve as a base for the hardware selection for the target context

2.1 The Source and Target Context

The source context serves as a base for the development process. It describes relevant business models, user groups and useful application concepts that can be adapted to the target context. The target context on the other hand includes a set of use cases that benefit from adopting these concepts and models for the development of services for the user group. The dimensions of the target context in conclusion form requirements to a specific hardware device and are crucial for the hardware selection.

The target context, a social terminal utilized in the scenario of a public conference, can be considers as a “mixed reality” (MR): The setting of a conference embedding “social media” is “augmented” by its social space and by adding virtual parts and information systems on displays to the scene. Interaction does not only take place physically any more but partly in virtual spaces. For a description of “virtuality” see [3]. It heavily depends on the application if it tends to be an Augmented Reality System or Ubiquitous Computing. Following this, a lot of research findings on MR systems apply to systems including types of public terminals [4], [5].

The chance for users to get accustomed to the use of a system also depends on if it was installed for long-term use or for an individual event. Placing the display in an appropriate position must take architectural conditions and the “people flow” into account to ensure findability and accessibility. The boundaries of “social space” and “physical interaction space” can be more or less sharp, depending on the application. Here, an information architecture as described in [6] can help users to apply their existing knowledge and mental models to the target context. It depends on the type of application whether content mapping has to be complete and symmetrical, or if a lightweight, partial view or asymmetric presentation is enough for the purpose.

Comparing the two contexts in our example, the social semantic web offers a range of valuable services and features that can be reused and adopted to the target context, e.g. typical planning, administration and communication services of community platforms to manage user profiles and access rights, features to store, share and publish content, can easily be adapted to fit the requirements of a system used during conferences. These services can be combined with the advantages of public terminals in the target context regarding the collaborative environment, the capabilities of presenting and sharing information.

2.2 The Business Models

Both source and target context base upon specific business models. They can be similar but commonly follow different objectives because of differences in the field of application and user group. Whether or not a source business model can be adapted to the target context, it is essential to define the target business model, as the further development process is based upon it. These business concepts in a conference scenario can include: building user or company networks, advertisement, information exchange for the participants´ benefit, participant inclusion in crowd sourcing concepts etc. Especially concepts for user inclusion and content creation form an essential part of the social semantic web and should be considered for the target business model.

2.3 The User Groups

Although the user groups of social media platforms and public terminals differ, they share common needs and expectations towards the interface. In both contexts there is an essential need to handle user profiles and access rights. If the services heavily base on the user profiles – like in communities where individuals or companies present themselves – the requirement expands to a management of various digital identities connected to different virtual contexts or real situations. The digital.me project funded by the EC, targets the development of userware for managing such personal information spheres. This includes integration of various social services as well as semantic search and browsing [7].

In both contexts in our example, the user benefits from features to personalize content that allow creating recognizable profiles and content representations in the conference context. From a users´ perspective, common requirements deal with security of and trust in private/public data, as well as trust in other users and their user-generated content. In both contexts they usually create and organise private data, as well as can use features to declare it as public data. On the one hand, for private data there is the need for specific security concepts to be able to define different security levels as on social communities. This is essential for both contexts including web and multi-user applications. On the other hand, public data must be identifiable as trustful or relevant regarding the content. If we have a look at the social semantic web this requirement is accomplished with trust building services that enable to rate the quality of content and trust in users.

Regarding the users´ interaction, it is crucial to design terminals in a way that makes it easy for passersby to recognize which services it offers at a minimum effort. The user has its own interests and goals, so the advantage of or incentives to start interacting with the terminal must to be communicated clearly, even from a distance. These characteristics in using public terminals are similar to people “surfing” the web without a specific target, but passing by different social media platforms by using following external links.

Reasons that inhibit people from interacting with public terminals are discussed more in detail in [5]. They range from not recognising the relevance of interacting to feelings of embarrassment, because it is not clear to them how much time it will cost, or they fear to look foolish. To overcome these fears, constant encouragement and demonstration is one mean giving potential users the ability to grasp what the interface does, e.g. by looking over the shoulder of others interacting with it. If the application concept is already familiar to users, they are able to build a mental model of what the application does much faster. Meaning, by using well-known application concepts of social media the user is able to interact with the terminal application with less effort.

2.4 Application Features and Services (Best Practices)

Regarding application features and services, the social semantic web comes up with a wide set of valuable concepts which can be adapted to contexts in which collaboration and multi-user interaction is an issue. This set includes services for “social tagging” as described in [8], where a user is able to add meta data to user generated content and other users for classification. In a conference context this is e.g. helpful to classify user-generated content like presentations on the one hand, but also to define experts and contact persons of a specific domain on the other hand.

“Crowd sourcing” concepts and related “open innovation” processes [9], enable the users to create content and bring in own ideas. For most web communities, crowd sourcing forms an essential part of the business concept because the mass of users has an immense power to innovate and solve problems collaborative. Similar for public conferences, a conference is often only successful and valuable for its participants if there is an agile contribution and services to share it.

At a glance over the most typical features and services of social media, the target context benefits from:

- Administration Services (user profiles, handling access rights, content management and life-cycle)
- Planning Services (e.g. to organize the conference visit, schedule meetings between participants and plan workshops)
- Communication Services (comments to published content or posts, messaging, video conferencing)
- Basic Services (for protocols, data transfer and navigation)
- Trust Building Services (to rate the trustfulness of users or the quality of content)
- Social Tagging (for a categorization of content and users)

In addition, the target context benefits of location-based services which uses the actual position of a social terminal, e.g. to schedule meeting at a specific place, in contrast to meeting in a specific virtual place in online communities.

2.5 The Hardware Selection

As hardware is a minor issue in the source context of the social semantic web, the selection of the appropriate hardware type and technology is only related to the target context. For this reason, hardware is chosen according to the application concepts and interaction design of the target context. However in real projects, restrictions due to existing hardware or to budget may influence the choice. As the hardware is crucial to the user experience and performance, the selection should be made carefully. This includes having in mind that in every single interaction step where a user has to “do something in a certain way just because the hardware forces the user to do so”, the overall acceptance of the system is suffering from it.

The form of interaction depends on a set of factors: the number of interacting users as we consider multi-user systems, the number of displays in one place, the scaling if it is a multi-user or a one-person-at-a-time setup. The range goes from monitor size to projection walls, horizontal interactive tables or tablets, to vertical boards or walls visible for a whole audience. For reasons of interaction design and user encouragement, display and input device can be separated like described in the Opinionizer [4]. For direct collaboration they should be an entity like in most multi-touch tables. Beside that, a system can allow for remote interaction. In this setup, mechanisms to identify users can require additional hardware features like cameras to detect markers and codes or reader devices for RFID and NFC cards.

Also the location of a hardware device demand varying requirements to the systems´ stability and encapsulation. A stationary terminal installed in an outdoor environment for example should follow industry standards regarding protection against weather conditions and vandalism.

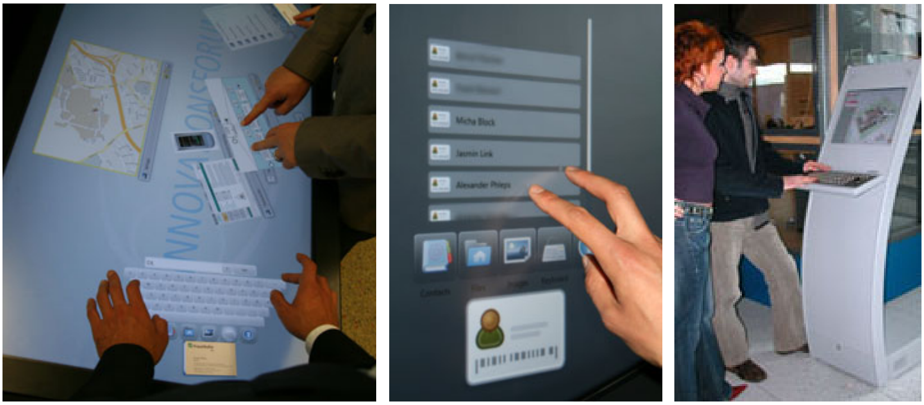


Fig. 2. Left, middle photo: GADjet Terminal which is included in the Innovationsforum at GAD is used for collaborative workshops for “contact and appointment management, collaborative innovation scenarios, presentation and discussion of content” [10]. Right photo: Information and Room Finding System Horst [11].

3 The Step Based Development Process

This section describes the step-by-step process of the context centric approach, in this case for a successful specification and design of social media application for public terminals. Following the process shown in Fig. 3., this section will detail each process step and discuss results and sources accordingly.

Before starting with the development of any application, it is essential to clearly understand its business concepts first. Therefore, the process starts with the specification of business concepts for the target context. As described in the previous section, the resulting document details the business model and implicitly refers to the main features of the application which can be used as services later on by the user. Further analysis of the business concept in the second step reveals a more explicit view on the required features, collected in the application concept. In parallel, target uses cases get assembled and documented. With every business model, the users build the key factors that are intended to use it. To gather them, the business concept is analyzed by categorizing user groups that are modeled in a third document. The three documents “Application Concept”, “General Use-Cases” and “User Groups” describe the target context for the application.

Knowing about our target context, in the third step, relevant source contexts get analyzed to find appropriate solutions for creating a mash-up application. In this example the source context is the social semantic web which its rich services for the users. Extensive research on the state of the art results in a list of appropriate social media applications (services and features) that could be integrated into the target application. For each source application a description of features and services provided gets documented in the corresponding “Source Application Concept”. Additionally, best practices are collected to support the design decisions on an architectural level, as well as for further reference within the implementation process. Finally, for all source application platforms the assigned community is analyzed and “Source

User Groups" get categorized. Categories in "Source User Groups" should be compatible with categories in the user groups identified for the target.

In the fourth step the mash-up architecture of the target application gets specified. Based on the collection of application concepts and user groups of the source context, the appropriate social media application services are selected that are to be integrated. The collected best practices of the source context help to sort and evaluate the various options. However, the decisions which social media applications or platforms will be supported is of high strategic relevance and should be considered with care. When comparing the two criteria: "services provided" and "fitting user groups", the necessary condition is a platform supporting the required services. While accessing, the fitting kind of users would be a sufficient condition. In the best case, both aspects are supported. When mixing services of different social media provider, the quality of the target application will depend on whether the combination of these services can be handled easily, or whether expensive mappings and transformations are required for information exchange between the services. These considerations should also incorporate whether the used concepts (models) are compatible for the user, e.g. one service might model a group as a set of persons, while a group in another services could refer to a discussion board.

Once the decision is made and the services have been selected, this results in a document describing the integration of the services as well as their (joint) functionality. This document contains a top-level architecture describing the inter-action between services and application. Additionally, it contains references to the service APIs intended to use with a description about their functionality. In parallel the specific use-cases get elaborated on a level of granularity that fit to the architectural approach.

Having defined the architecture and use-cases, the fifth step elaborates the concept and design of a user interface. The goal is to develop a user interface of high quality considering the target hardware device and a good usability in the target context, supporting the application of the business concept. In this case the user interface development focuses on the usage of a public terminal.

While the actual hardware of the platform is still undefined regarding its technology and construction at this step, already some general features and properties of the platform are known. They are collected and available as interaction patterns and design patterns [12] that are specific for this kind of devices. Additional features and requirements on the one hand depend on the target context, the usage and the stationary location of the terminal. On the other hand, they are derived from the business concept and description of the use-cases. Based on this information, the user interface concept is elaborated and tailored to the identified user groups. In a story board, sequences of views specify its behavior. Eventually, an analysis of the UI concept provides detailed requirements for the appropriate hardware of the public social terminal.

A collection of hardware specifications of different touch technologies available on the market for public terminals (described in [13]) helps in the sixth step to decide which technology and construction serves best in the specific context, based on the identified requirements. This sixth step accomplishes the general specification phase of the social media application. In the following implementation cycle the actual application gets developed and tested. The evaluation of each iteration cycle is done in a typical user-centric approach and with inclusion of the selected design and interaction patterns.

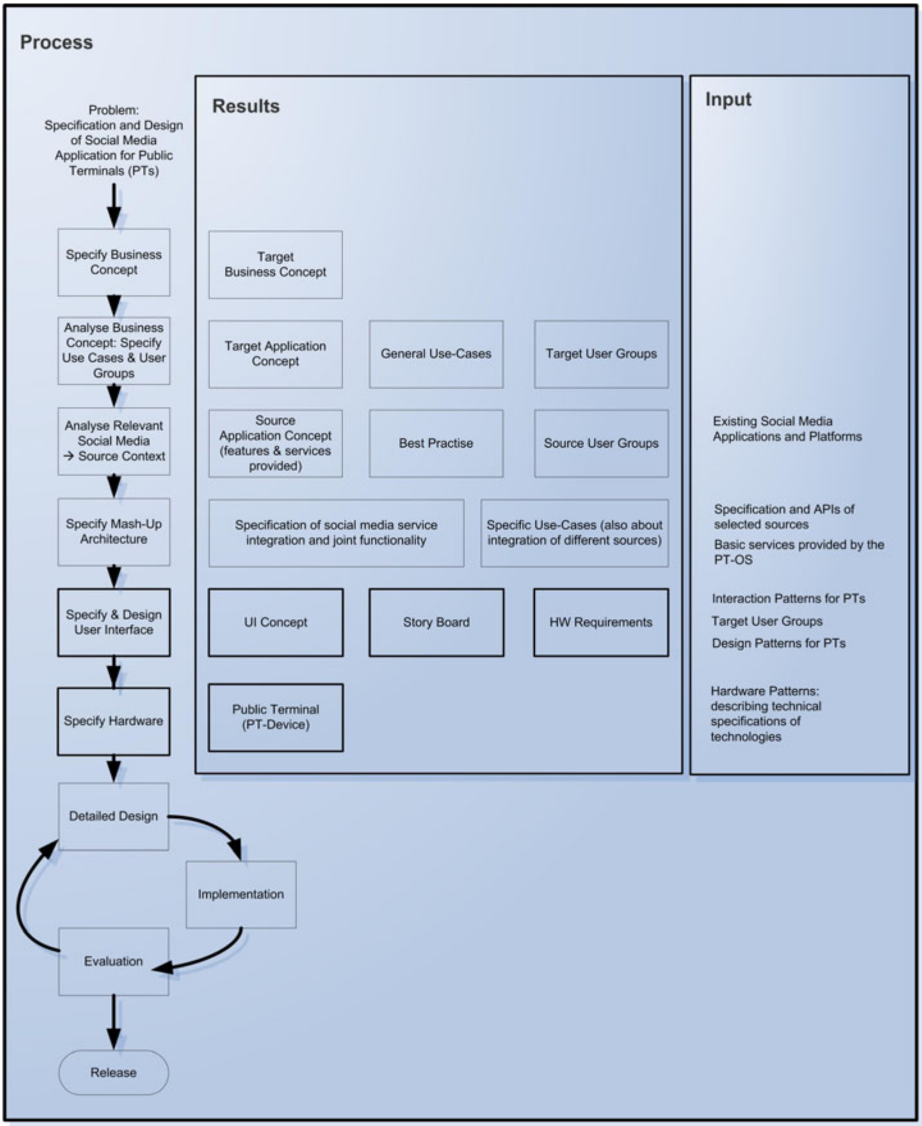


Fig. 3. Step-by-step development process

4 Using the Context Centric Approach for Evaluation

The process described in the previous section focuses on the development of interactive systems by adapting concepts of a source context to a target context. As it uses design and interaction patterns for a specific context, as well as hardware specifications of hardware devices, it can also be utilized for an evaluation of existing concepts and interactive systems. At that stage, the target context with its business models, user groups and application concepts already exists. For the evaluation, single application features and concepts are compared with appropriate design and interaction patterns to check if there is room for improvement. On hardware level, the collection of specifications provides comparable attributes to identify if there is a technology available which fits better to the context and its related application concept.

As a key benefit, this approach enables to effectively adopt key application concepts and features to be able to apply them in different contexts and on different hardware devices. On an abstract layer, this approach can be applied to any source and target context by changing the specific "input" patterns during the specification phases. Thus, it provides a powerful and flexible approach for a context based development and evaluation of interactive systems.

References

1. Apple HIG, http://developer.apple.com/library/ios/#documentation/UserExperience/Conceptual/MobileHIG/Introduction/Introduction.html#//apple_ref/doc/uid/TP40006556; http://developer.apple.com/library/ios/#documentation/General/Conceptual/iPadHIG/Introduction/Introduction.html (last checked October 21, 2010)
2. Microsoft Surface Framework Guidelines, http://msdn.microsoft.com/en-us/library/ff318692%28v=Surface.10%29.aspx (last checked October 21, 2010)
3. Milgram, P., Takemura, H., Utsumi, A., Kishino, F.: Augmented Reality: A class of displays on the reality-virtuality continuum. In: Proceedings of Telemanipulator and Telepresence Technologies. SPIE, vol. 2351-34 (1994)
4. Schnädelbach, H., Galani, A., Flintham, M.: Embedded Mixed Reality Environments. In: Dubois, E., Gray, P., Nigay, L. (eds.) The Engineering of Mixed Reality Systems. Springer, Heidelberg (2010)
5. Brignull, H., Rogers, Y.: Enticing People to Interact with Large Public Displays in Public Spaces. In: Rauterberg, M., et al. (eds.) Human-Computer Interaction INTERACT 2003. IOS Press, Amsterdam (2003)
6. Resmini, A., Rosati, L.: The Semantic Environment: Heuristics for a Cross-Context Human-Information Model. In: Dubois, E., Gray, P., Nigay, L. (eds.) The Engineering of Mixed Reality Systems. Springer-Verlag London Limited, Heidelberg (2010)
7. Project digital.me, funded by the EC, http://www.dime-project.eu/ (last checked February 10, 2011)
8. Vander Wal, T.: Folksonomy::vanderwal.net, http://vanderwal.net/folksonomy.html (last checked June 1, 2007)
9. Mast, C., Zerfaß, A.: Neue Ideen erfolgreich durchsetzen. Das Handbuch der Innovationskommunikation. pp. 20–21. Frankfurter Allg. Buch. Frankfurt a.M (2005)
10. GAD Innovationsforum: Münster, Germany, http://www.gad.de/veranstaltungen2/innovationsforum.html (last checked February 10, 2011)
11. Information System Horst at Hochschule der Medien Stuttgart, http://www.hdm-stuttgart.de/horst (last checked February 10, 2011)
12. Block, M., Link, J., Phleps, A.: Design Patterns for Multitouch-TUIs. Engineering Patterns for Multi-Touch Interfaces 2010. In: Workshop of the ACM SIGCHI Symposium on Engineering Interactive Computing Systems, Berlin (2010)
13. Block, M., Laufs, U., Link, J., Ardilio, A., Schuller, A., Bierkandt, J.: Studie Multi-Touch. Technologie, Hard-/Software und deren Anwendungsszenarien. In: Spath, D., Weisbecker, A. (eds.) Fraunhofer IAO, Stuttgart, pp. 3–42 (2009), https://shop.iao.fraunhofer.de/details.php?id=437 (last checked February 10, 2011)

Accessibility for Older Users through Adaptive Interfaces: Opportunities, Challenges and Achievements*

Rob Edlin-White[1], Sue Cobb[1], Mirabelle D'Cruz[1], Anne Floyde[1], Sarah Lewthwaite[2], and Johann Riedel[3]

[1] Human Factors Research Group, The University of Nottingham, University Park, Nottingham NG7 2RD, United Kingdom
[2] Learning Sciences Research Institute, School of Education, The University of Nottingham, University Park, Nottingham NG7 2RD, United Kingdom
[3] Business School, The University of Nottingham, University Park, Nottingham NG7 2RD, United Kingdom
{epxrwe,sue.cobb,mirabelle.dcruz,anne.floyde,ttxsem, johann.riedel}@nottingham.ac.uk

Abstract. It has been widely suggested that accessibility to technology for older people could be improved by the provision of adaptive user interfaces. However there has been little practical work in the area. The MyUI[1] project sets out to explore this area and to build infrastructure and three demonstrator applications based on interactive TV technology. This paper, looking at the project from a Human Factors perspective, explores whether accessibility for older people through dynamically adapting interfaces is a realistic goal, identifies some challenges and research questions, and provides an insight into some achievements and ongoing work in the project.

Keywords: "older people" "adaptive interfaces" "accessibility".

1 Introduction

Systems which adapt their user interface or content dynamically are increasingly deployed for various purposes including personalised learning [1], increasing sales on ecommerce sites [2] and user productivity [3].

For over a decade, researchers have been suggesting that adaptive interfaces could be a promising vehicle to deliver universal accessibility, to open up access to an increasingly digital society to users with physical, sensory or cognitive impairments [4,5,6,7,8].

The need for more accessible technology for work, leisure and health grows ever more urgent. Societies are aging; people are living for longer with age-related impairments; retirement from economic work is being postponed, and more and more services are available only or mainly through technology.

* This work was funded by the European Union MyUI project under grant FP7-ICT-248606.
[1] http://www.myui.eu

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 483–489, 2011.

However there appears to be little research into exactly how automatic interface adaptation can support accessibility. The area demands scrutiny as adaptive interfaces and devices are becoming available and technology is increasingly able to detect user characteristics. Recent projects such as MyUI, GUIDE[2] and SUS-IT[3] are now beginning to explore this area.

This paper discusses the work being conducted within the MyUI project, specifically from the Human Factors perspective. The many technical achievements of the project are described elsewhere [23, 24, 25].

2 Aging Societies and Technology

Across the world, societies are aging, and therefore the prevalence of age-related impairments is increasing dramatically. Access to technology for older people can be limited by a range of impairments in the physio-motor, sensory and cognitive domains, and the prevalence and impact of these are well documented [e.g.9, 10, 16].

Some impairments can be mitigated with assistive devices or by other means; e.g. spectacles, brighter lighting, reminders (for prospective memory), but the benefits are limited and the need for more accessible technology remains.

3 An Opportunity for Adaptive Systems

Technology is increasingly necessary to access basic services in an increasingly digital society [17] and also has the potential to deliver services of specific value to older people in their domestic environments.

Technological standards and guidelines concerning general application design for older people and those with disabilities are particularly developed in web accessibility discourse [e.g. 11, 12, 13, 14]. Advances are being made in other areas too, e.g. television [15] .The ISO document "*Ergonomics Data and Guidelines for the application of ISO/IEC Guide 71 to products and services to address the needs of older persons and persons with disabilities*" [16] represents an excellent resource in this important and growing field.

Standards and guidelines of this sort, as well as inclusive design methodologies, have marked important steps towards accessibility. However progress towards universal accessibility has been limited in several respects.

There have been long standing aspirations for "Universal Design", or "Design for All"; products which anyone can use [17]. However for a variety of reasons the notion of a single universal interface seems not to be practical. Several researchers [4, 5, 6, 7, 8] have suggested adaptive interfaces are a plausible solution to the desire for universal accessibility, providing an affordable and effective design-once, adapt-to-all strategy.

Kelly et al conceive three waves of critique which shift the focus of accessibility away from a standards-based universal model, to critical, holistic and ultimately adaptable positions [18]. This review notes that early universal methods of design frequently overlook the importance of context in which technologies are used. Further to this, the authors observe a requirement to recognise the 'idiosyncratic needs and

[2] `http://www.guide-project.eu/index.php`
[3] `http://sus-it.lboro.ac.uk/`

preferences of individual users' and to cater for them accordingly. This moves away from a homogenised archetype to an accessibility model more accommodating to diversity. This is important given the contrary requirements for different impairments.

Kelly et al's conception of Adaptability recognises the active role that the user can play in achieving accessibility and designing their own experience [18]. However in some cases, customisation processes pose accessibility problems, so automatic adaptation has an attraction.

Automatic adaptation involves three stages [19]:

- "Afference" - Sensing user behaviours,
- "Inference" - Inferring information about user,
- "Efference" - Presenting the interface in a way customised for that user.

An adaptive system designed specifically to improve accessibility would therefore set out to sense user behaviours which might indicate an ability or impairment or accessibility need, and, from patterns of such observations, make inferences about the user to be stored in a user profile. When interacting with the user, the interaction design would be constructed or modified in real time to optimise accessibility for that specific user. It would need some method of identifying the user, so that the correct user profile is accessed.

The MyUI project sets out to build an infrastructure to support affordable development of accessible adaptive applications, and three demonstrator applications based on interactive TV technology.

4 Technical Challenges

An ideal adaptive system to enhance accessibility would need to instantly and accurately determines a user's accessibility needs and preferences without any overt intervention or obtrusive devices, and adapt immediately to provide an interaction style perfectly suited to that user. This seems an over-optimistic and unrealistic aspiration.

The proposed approach poses several significant questions currently under-represented in literature:

- What, if any, observable behaviours do older users exhibit when they encounter accessibility problems? Do all relevant impairments result in some observable behaviour? How consistently do they do this?
- What devices can detect and measure such behaviours, and how accurately?
- To what extent will older users tolerate extra devices and/or extra interactions which are intended to measure accessibility needs? Must a system rely purely on invisible ambient detection?
- How accurately is it possible to infer accessibility needs from observed behaviours? And to what extent does this require a number of observations over a timescale? An additional complexity arises from the fact that older people sometimes use assistive devices (such as hearing aids) erratically and inconsistently, creating erratic changes in accessibility needs.
- What features of an interface can be adapted in a way which older users will find acceptable and beneficial? Design standards for older people [e.g. 16] recommend stable, consistent interfaces, rather than changeable ones. Adaptations must be so

designed that the benefits of increased accessibility outweigh the disadvantages of changeability.
- When adapting an interface, to what extent should the user be advised or consulted on the proposed change?
- When a system gathers data about a user's abilities and limitations, what responsibilities does the system have if it discovers information which may indicate a significant medical condition? Do privacy rights prevail?

5 Methodological Challenges

Initiatives to improve accessibility should take an ethical stance on defining, understanding and measuring accessibility [20]. For many older people the affordability of the technology and availability of support may be key issues. An evaluation methodology which ignored such factors would not be ultimately in the user's interests. Users should be fully engaged throughout.

There are also some significant more general challenges, arising from working with older people with such varied abilities and limitations:

- It can be difficult to obtain and retain informed consent from users with possibly poor hearing, poor memories or comprehension difficulties.
- It can be hard for older people to envisage or imagine usability issues and the benefits and shortcomings of unfamiliar technologies. Eliciting requirements and discussing proposed solutions can be difficult unless there is some kind of prototype to provide more concrete experience.
- Older people sometimes have less focus on a task, and are more prone to distractions and digressions. Working with older people, especially in their own homes, can sometimes involve quite a lot of time spent on digressions and "social niceties" [11]. This increases research timescales.
- Bringing older people into a lab situation can be problematical, because of heightened artificiality of the environment, and because of matters relating to the "duty of care". This means results from lab studies may be subject to distortion and have little correspondence to their real lives.

6 Some Achievements within the MyUI Project

A series of Focus Groups with older people were conducted in field locations in the UK and Spain. The groups were stimulated by handing round a number of devices and visual aids, asking questions, mingling and conversing, demonstrating and facilitating group conversations. Researchers used a plenary session at the end to check whether anecdotal observations were common experience or specific to individuals. This process provided a wealth of rich qualitative feedback, rooted in practical experience. The researchers gained an in-depth and ecologically valid understanding of the users, their environments, their objectives, their existing use of relevant technologies and hopes, difficulties, preferences and achievements in their use of technology.

This information was supplemented by a series of interviews with therapists, carers, professionals in care of older people, wardens, Day Centre staff etc. All research was informed by prior literature study.

From this research base, the project produced:

- Personas – archetypal representations of some possible users;
- Scenarios – descriptions of possible situations in which a proposed technology could be used;
- A Requirements document

The Requirements document [21] arose from study of existing literature as well as the research. It catalogued User Characteristics, User Requirements (including Accessibility Requirements), Functional Requirements, Technical Requirements, Developer Requirements and Non-functional Requirements.

There were remarkable contrasts in the lifestyles, abilities, patterns of aging and therefore requirements between users in the UK and Spain [22].

It proved difficult to elicit and verify requirements specifically for adaptation as the older people had little experience of adaptive systems, the devices handed round in Focus Groups had few adaptive features, and it proved difficult to get participants to imagine such systems from verbal descriptions.

A later task involving formative evaluations is in progress. This consists of studies at three levels, as illustrated in Figure 1.

In this pattern, more formal (Level 1) studies of specific human characteristics or components of possible technical solutions precede and inform the design of more complex components. These studies will thus make novel contributions to scientific knowledge. They are more abstract, involve more rigorous controls, tighter participant selection criteria, more specific participant instructions and more objective measurements, and tend to use simpler interfaces to minimise any potentially distracting features.

Higher level studies will use more complex and realistic prototypes, increasing realism and ecological validity but also at the risk of reduced controls and higher potential for participant distraction.

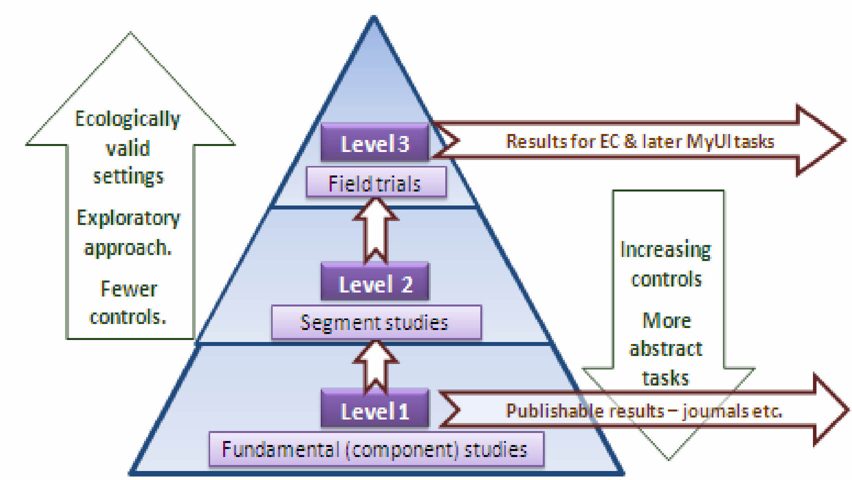


Fig. 1. Study Structure

At Level 3, potential demonstrator products will be tested, and for these purposes, longer term, ecologically valid field-based studies are appropriate.

7 Conclusion

Adaptive user interfaces are widely prescribed as a method to make technology more accessible to older people with age-related impairments. There are, however, a huge number of research questions which need to be addressed in order to make progress in this area. The MyUI project is providing promising insights and progress towards this. We are currently conducting level 1 studies; e.g. to identify what if any observable user behaviours are exhibited when a font size on a display is too small. Further work will conduct higher level studies including eventual evaluations of the MyUI demonstrator applications.

Acknowledgements. The authors acknowledge the help of other partners in the MyUI consortium in this work: Phillips Consumer Lifestyle, ForschungszentrumInformatik, CleverCherry.com, ISOIN, ADG, SOTE, Universidad Carlos III de Madrid, Birmingham City Council, Getafe City Council, and especially the project manager, Matthias Peissner and his colleagues at Fraunhoffer IAO.

The authors are also grateful for the help and cooperation of numerous older people, carers and professionals in care of older people for their involvement in Focus Groups, interviews and other activities to support this work.

References

1. ISO/IEC 24751-1 Information technology –Individualized adaptability and accessibility in elearning, education and training – Part 1: Framework and reference model. ISO/IEC 24751-1:2008. Treviranus, J., Nevile, L., Heath, A (eds.), http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=41521
2. Zhang, X., Edwards, J., Harding, J.: Personalized online sales using web usage data mining. Computers in Industry 58, 772–782 (2007)
3. Hartmann, M.: Challenges in Developing User-Adaptive Intelligent User Interfaces. In: Proceedings of the 17th Workshop on Adaptivity and User Modeling in Interactive Systems, Darmstadt, Germany, pp. 6–11 (2009)
4. Fisk, A.D., Rogers, W.A.: Handbook of human factors and the older adult. Academic Press, London (1997)
5. Stephanidis, C., Salvendy, G., Akoumianakis, D., Bevan, N., Brewer, J., Emiliani, P.L., Galetsas, A., Haataja, S., Iakovidis, I., Jacko, J., Jenkins, P., Karshmer, A., Korn, P., Marcus, A., Murphy, H., Stary, C., Vanderheiden, G., Weber, G., Ziegler, J.: Toward an Information Society for All: An International R&D Agenda. International Journal of Human-Computer Interaction 10(2), 107–134 (1998)
6. Stephanidis, C., Emiliani, P.L.: Connecting to the Information Society: a European Perspective. Technology and Disability Journal 10(1), 21–44 (1999)
7. Gregor, P., Newell, A.F.: Designing for dynamic diversity - making accessible interfaces for older people. In: ACM, Jorge, J., Heller, R., Guedj, R. (eds.) Proceedings of WUAUC 2001 (2001 EC/NSF Workshop on Universal Accessibility of Ubiquitous Computing: Providing for the Elderly, Portugal, May 22-25, pp. 90–92 (2001)

8. Stephanidis, C.: User interfaces for all: New perspectives into human-computer interaction. In: Stephanidis, C. (ed.) User Interfaces for All, pp. 3–17. Lawrence Erlbaum, Mahwah (2001)
9. Smith, S., Norris, B., Peebles, L.: Older Adultdata: the handbook of measurements and capabilities of the older adult – data for design safety. UK Department of Trade and Industry, London (2000)
10. Fisk, A.D., Rogers, W.A., Charness, N., Czaja, S.J., Sharit, J.: Designing for older adults: Principles and creative human factors approaches. CRC Press, London (2004)
11. Pernice, K., Nielsen, J.: Web usability for senior citizens. Design guidelines based on usability studies with people age 65 and older. Nielsen Norman Group, Fremont (2002)
12. Australian Human Rights Commission, World Wide Web Access: Disability Discrimination Act Advisory Notes Version 3.3.1, Australian Human Rights Commission (2009), `http://www.hreoc.gov.au/disability_rights/standards/www_3/www_3.html` (accessed October 10, 2010)
13. Caldwell, B., Chisholm, W., Vanderheiden, G., White, J.: WCAG 2.0: W3C Working Draft, W3C (November 19, 2004), `http://www.w3.org/TR/WCAG20/` (accessed July 20, 2010)
14. Caldwell, B., Cooper, M., Reid, L.G., Vanderheiden, G.: Web Content Accessibility Guidelines (WCAG) 2.0. W3C (2008), `http://www.w3.org/TR/WCAG20/`
15. Carmichael, A.R.: Style Guide for the Design of Interactive Television Services for Elderly Viewers. Independent Television Commission, Standards and Technology. Kings Worthy Court, Winchester (1999)
16. ISO/TR22411:2008: Ergonomics data and guidelines for the application of ISO/IEC Guide 71 to products and services to address the needs of older persons and persons with disabilities, British Standards Institute (2008)
17. Emiliani, P.L., Stephanidis, C.: Universal access to ambient intelligence environments: Opportunities and challenges for people with disabilities. IBM Systems Journal 44(3), 65 (2005)
18. Kelly, B., Nevile, L., Sloan, D., Fanou, S., Ellison, R., Herrod, L.: From Web Accessibility to Web Adaptability. Disability and Rehability: Assistive Technology 4(4), 212–226 (2009)
19. Paramythis, A., Weibelzahl, S.: A Decomposition Model for the Layered Evaluation of Interactive Adaptive Systems. In: Ardissono, L., Brna, P., Mitrović, A. (eds.) UM 2005. LNCS (LNAI), vol. 3538, pp. 438–442. Springer, Heidelberg (2005)
20. Picking, R., Robinet, A., Grout, V., McGinn, J., Roy, A., Ellis, S., Oram, D.: A Case Study Using a Methodological Approach to Developing User Interfaces for Elderly and Disabled People. The Computer Journal 53(6), 842–859 (2009)
21. MyUI Consortium (2010), MyUI: Mainstreaming Accessibility through Synergistic User Modelling and Adaptability Deliverable D2.1 Requirements for User Interface Adaptation, `http://www.myui.eu/index.php?option=com_remository&Itemid=189&func=fileinfo&id=5`
22. Edlin-White, R.W., Floyde, A., D'Cruz, M., Cobb, S., Riedel, J.C.K.H.: User interface adaptation for accessibility by older users and the challenges of cultural diversity. Presented at BCS HCI Conference, Dundee (September 2010)
23. Peissner, M., Schuller, A., Spath, D.: A Design Patterns Approach to Adaptive User Interfaces for Users with Special Needs. In Preparation; to be Published in these Proceedings (2011)
24. Gacimartin, C., Hernandez, J.A., Larrabeiti, D.: A middleware architecture for designing TV-based adapted applications for the elderly. In: Preparation to be Published in these Proceedings (2011)
25. Takacs, B., Simon, L.: Sensing User Needs: Recognition Technologies and User Models for Adaptive User Interfaces. In Preparation; to be Published in these Proceedings (2011)

Computer Usage and User Experience in Jordan: Development and Application of the Diamond Model of Territorial Factors

Fuad EL-Qirem[1] and Gilbert Cockton[2]

[1] Al-Zaytoonah University of Jordan, Department of Software Engineering, Amman, Jordan
[2] Northumbria University, School of Design, Newcastle, United Kingdom
fqirem@yahoo.com, gilbert.cockton@northumbria.ac.uk

Abstract. The Diamond Model structures territorial factors of relevance to Interaction Design into five segments with objective and subjective facets. The initial structure and content was derived iteratively on the basis of a literature survey. Three field studies were then used to populate the model with Jordanian instances of cultural and other territorial variables, and to add new variables to the model. The model can act as design resource that combines concerns from HCI4D (HCI for Development) with cultural variables that extend beyond the acceptability and suitability of user interface elements. The derivation and evolution of a Jordanian instance of the model is presented, with brief suggestions on how it could be used by software development teams.

Keywords: HCI4D, Diamond model, cultural differences, geographical differences.

1 Introduction

Interaction Design, supported by HCI research, has to balance and integrate choices of software and media features with understandings of users and other stakeholders. Stakeholders vary in their capabilities, aspirations, aversions, needs, wants, and preferences, interpretations, to highlight only a few forms of differences among people. While many differences are very individual and idiosyncratic, we tend to design for groups of people, rather than for individuals. We design for what stakeholders have in common. Such similarities are often cultural, that is, they reflect the social environments that shape phenomena such as motives, values, meanings, obligations and expectations. However, not all differences are cultural. Economic, political and geographic factors also influence the suitability and acceptability of designs. The field of HCI4D (HCI for Development, Chetty and Grinter 2007) contrasts with earlier work on *culturability* (Barber and Badre 1998), by taking a broad view of the territorial factors that influence the suitability and acceptability of interactive systems. Culturability combines usability and cultural sensitivities to design user interfaces that are appropriate and acceptable for specific cultures. However, this may not be all that is required to successfully localize interactive software for specific geographic territories.

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 490–499, 2011.

Target markets for software design are often territorial. This is much more the case with software localization, although often language is the main focus here (e.g., for Francophone, Hispanic or Arabic markets). A focus on culture alone is thus inadequate. HCI4D in particular needs to consider economic, political and geographic factors to make software, for example, affordable for buyers, compatible with government policies, and suitable for the climate, country/region size, and technological infrastructure. A territorial focus also makes it possible to consider the range of cultures within a country or region, e.g., youth and organizational cultures.

A further advantage of considering a broad range of territorial factors is that national differences do not all need to be attributed to culture. Discussions of national culture often arouse concerns about stereotyping, overgeneralization, bias and even racism. Consideration of a broad range of territorial factors reduces concerns here. Even so, culture remains important, even though it has been defined in many ways in different disciplines. Within anthropology, Kluckhohn's definition has been influential. Kluckhohn (1951) defined culture as consisting in patterned ways of thinking, feeling and reacting, acquired and transmitted mainly by symbols, constituting the distinctive achievements of human groups, including their embodiments in artefacts; the essential core of culture consists of traditional ideas and especially their attached values. Such patterns are relevant to Interaction Design, because users act, evaluate, feel, think, and achieve when using computers. The central role of values in culture indicates that software features and usage could draw positive or negative reactions according to users' culture. While economic, political and physical geographic factors are often emphasised in HCI4D, the values shaped by cultural variables also shape suitability and acceptability.

Therefore, designers of new technologies or localizers of existing ones need to consider cultural and territorial factors. Currently, when designing for developing countries, designers need to combine findings from HCI4D with culturability research. To reduce the effort needed here, we have developed a new model called the "Diamond Model" which can guide designers and developers to consider how users may think, react, behave when using a design, as well as the economic and other resources and constraints that apply, especially in developing countries.

The Diamond Model was developed in four stages. A literature review of culturability research and general sources on culture provided the bases through a process of iteration for the model's structure of five segments of variable groups. We then added instances of variables through three studies. In the first initial exploratory study, interviews with IT support and developers highlighted the more common types of IT usage problems. We chose IT support experts for this pilot study because they have extensive experience of problems that face Jordanian users, were easy to access, and could offer a rapid overview of usage problems.

The second study repeated Lazar and colleagues' (2006) US study in Jordan, and compared their results with results from Jordan to identify differences between the two countries in levels and patterns of frustration and anxiety. Lazar and colleagues' (2006) study was published as our initial exploratory study completed. Repeating it provided a good opportunity to explore reactions to computer usage in Jordan, since Lazar and colleagues' reports appeared to be very different to the situations reported in our first study. We could explore if frustration levels differed between the US and Jordan and, if so (as expected), we could then explore reasons for them through triangulation interviews. The aim of our second study was thus not to show that we would get different frequencies to Lazar and colleagues, but to use any differences here as a basis for exploring whether Jordanian users believed any of the revealed differences, how they evaluated these differences, and most importantly, how they explained them. Thoughts and feelings here would be shaped by culture, but there may also be economic or political explanations for differences in user experience between samples from two countries.

The third study studied computer usage experiences in Jordan through semi-structured interviews in the work place, which included open questions. The interviews had three parts: the first part was about the current software programs that they used and the main purpose of using a computer at work, home, and in public places such as internet cafés. The second part explored what participants thought their future computer should be and what they need from it: what type of changes do users need? The third part asked for suggestions as to what motivates them to use computers. For example, what could make computers easier for use? As part of this, participants were also asked to give their explanations for causes of computer problems in Jordan identified in previous studies.

All three studies provided Jordanian instances of cultural variables in the initial Diamond Model; however, they also revealed new cultural and other territorial variables that had not been considered in the literature in 2005. Many of these new variables were related to how users evaluated their usage experience, rather than the main focus within culturability research on the acceptability of user interface design elements. Also, for some cultural variables that are stressed in the literature, no Jordanian instances were found in any of our three studies.

2 The Diamond Model

Four meta-models of culture guided early research on software localisation (Del Galdo and Nielsen 1996). In theory, meta-models make it easier to compare cultures, but each of the four is an inadequate foundation for design work. Hofstede's (1980) *Pyramid Model* locates culture between individuals and nature, and is little more than a diagrammatic definition. The *Objective and Subjective Model* (Triandis 1972) provides a useful distinction between observable objective and often persistent manifestations of culture (Kluckhohn's *transmitting* symbols) and an implicit subjective core (Kluckhohn's *transmitted* values). The *Onion Model* (Trompenaars 1993) uses a physical analogy to relate an objective outer layer of cultural artefacts to inner layers of subjective norms, values and longstanding environmental adaptations. The *Iceberg Model* (Victor 1992) uses an alternative physical analogy to contrast an objective visible surface with subsurface subjective unspoken and unconscious rules. Both physical analogies are inappropriate as the subjective components are too opaque or occluded. Neither captures the dynamic relationships between objective and subject aspects of culture. A more appropriate physical metaphor appears to be a *diamond*.

Like onions and icebergs, diamonds have a visible upper surface (over their *crown*) and an occluded lower part (the *pavilion*). However, unlike an opaque onion, light emits from the crown, having entered via it, and then been reflected within the facets of the pavilion. This better captures the dynamics of the relationships between objective and subjective aspects of culture. Light takes paths that pass through all parts of the diamond.

The diamond metaphor was extended to include vertical *segments* that could group related objective and subjective cultural variables. Figure 1 shows the relation between variables and their objective and subjective sub-segments. The literature survey identified five large segments of culture. Each sub-segment contains groups, and each group contains specific variables, for which actual instances will vary across cultures. Each group could impact both design preferences and how usage is evaluated. Some example groups are labelled in Figure 1. For example; semiotics and language in Jordan reflect the Arabic language and its Jordanian dialect. Arrows show light paths through the diamond, e.g. from thought and feelings to social interaction, or via semiotics to language.

In Figure 1, the top (crown) of the diamond contains objective variables which are observable and the bottom (pavilion) contains subjective variables which become deeper and decreasingly obvious. Segments are ordered from the left from the most obvious to the more subtle. It may help here to think of light coming from the left, and thus more strongly illuminating the leftmost segments, which are more obvious due to the nature and extent of their objective sub-segment, e.g., material culture is explicit and obvious to any visitor, as is language. Politics, economics and social behaviours are progressively less obvious. A visitor may not even notice a social behaviour, still less be able to understand its underlying thoughts and feelings. The five segments of territorial factors in the Diamond Model are now briefly described and discussed.

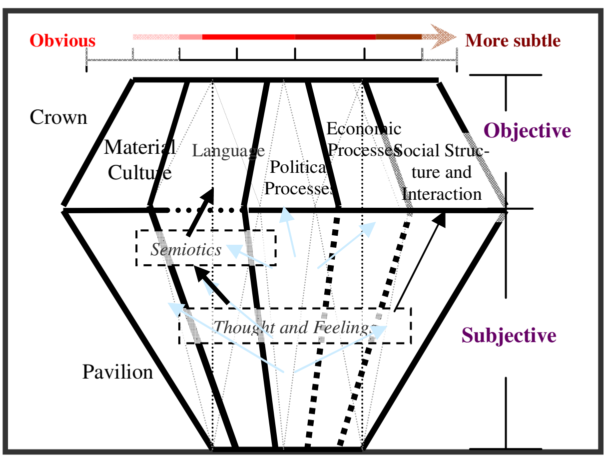


Fig. 1. The diamond model with five segments

2.1 Material Culture

Material culture covers the most obvious objective variables. In the crown of the diamond, they are instantly recognizable and easy to encounter. Material culture also includes media such as newspapers and television, as well as digital artefacts. Its subjective sub-segment is largely concerned with symbolism for example; the culturally preferred meanings and qualities of artefacts. Much of this overlaps with the semiotics of spoken and written language; hence the semiotics group of subjective cultural variables span two sub-segments. Material culture includes arts, building, houses and monuments, crafts and decorative arts, food, and literature.

2.2 Language and Semiotics

Language gives rise to objective variables such as vocabulary, grammar, dialect and accent, accessible and clear in the crown of the diamond, but semiotics are subjective variables in the pavilion and as such are deeper and harder to reach, and not immediately obvious to outsiders.

2.3 Political Processes

Some political variables are objective and thus easy to describe, such as political institutions and public authorities. The subjective sub-segment covers political attitudes, for example, to democracy, authority and use of violence. As a concrete example of variables within this segment, HCI research had already recognised the relevance of political contexts (Smith and Yetim 2004) and the political status of the English Language (Del Galdo & Nielsen 1996).

2.4 Economic Processes

Economic variables include wealth and material resources, which are objective and thus in the crown sub-segment. Subjective variables in the pavilion include conceptions of worth, value and affordability. These reflect consumer choices, and extend to free goods and services such as e-cards for birthdays etc. As such, economic processes overlap with social and political processes via the broadest subjective variable group of thoughts and feelings.

2.5 Social Structure and Interaction

Social interactions give rise to objective cultural variables that include observable behaviours in religion, social communication and family customs. All these variables will be in the crown, and thus easily recognizable. Thought and feelings are common to all subjective sub-segments, but they are most frequently associated with social structure and interaction in the literature on culture. We thus associated the social sub-segment of ‘thought and feelings’ with social structure and interaction but recognise that all subjective sub-segments overlap extensively.

2.6 Summary

Following the literature survey and finalisation of the structure of the Diamond Model, the segments were judged to be relevant to Interaction Design for developing countries in the following order, taking into account the relative attention to each within the literature on HCI and culture.

1. Language and semiotics.
2. Material culture.
3. Social structure and interaction.
4. Political processes.
5. Economic processes.

As we populated the model with Jordanian examples during our studies, we were interested whether this ordering would change on the basis of empirical evidence of frequency and impact of territorial variables from each of these five segments.

3 Extending and Populating a Jordanian Instance of the Diamond Model

All three studies added new territorial variables. For example, only two disjoint political variables had been noted in the existing HCI literature on culture and localisation, but the first study with IT experts identified the importance of government educational policy on IT literacy, and the provision of local community IT training centres. Five further social variables (age hierarchies, attitudes to work, religious beliefs, and IT literacy in specific job roles) were identified, as well as two economic variables (support for local currencies, availability of international products and services). Note that, for example, IT literacy in specific job roles is a cultural variable, and thus is relevant in different ways in different territories. There will be occupational differences in IT competence in all countries, but the distribution of expertise will not be the same in all of them.

The first study used semi-structured interviews with 19 IT support, and web developers and designers in Jordan from three sectors: education, telecommunication, and banking. Analysis of interviews was in three phases. The first quickly formed a descriptive overview of computer problems that face users in the university, telecommunications and banking. The second formed a list of common problems across all three sectors. The third phase constructed an affinity diagram of all the reported problems. The three phases of analysis were chosen to explore how the impact of cultural variables could be exposed and analysed.

The second study repeated an existing HCI study (Lazar et al. 2006), not to replicate it, but to use it as a *probe*. The number of participants in Jordan was 109 students, and 52 employees. In the US study were 107 students, and 50 employees. Lazar and colleagues' research used pre- and post-session affective questionnaires and an incident reporting form to study anxiety, frustration and related affective responses during computer usage. Differences in self reported anxiety and frustration (Qirem, Cockton and Loutfi 2007) between the US and Jordanian samples provided a focus for triangulation interviews, where these differences were discussed.

The third study was less focused than the first two, with the result that its open interviews and user participants (as opposed to experts) identified a wide range of new territorial variables for the Diamond Model. There were 24 participants (16 IT support and software developers, 8 non-IT people) from Al-Zaytoonah university of Jordan, all computer users. The data was analysed in detail by constructing an affinity diagram to support identification of instances of cultural variables and the new variables that can be added to the Diamond Model. Within the material culture segment, a specific variable on colour preferences was added (NB this is not the same as cultural differences in the *meanings* of colours). Within the language segment, a variable was added for the impact of overly direct translations, which although intelligible can be annoying. Within the political segment, three variables were added: corporate policies (e.g. on language use), government policies on language education, and trade sanctions. The last was identified through comments on Syria, where trade sanctions have resulted in a strong Arabic language software industry, in contrast to Jordan, which has close relations with the UK and no imposed trade embargo. Within the economic segment, a group of three variables were added for affordability of IT access, i.e., affordability of computers, internet access and training. Within the social segment, a specific variable was added on gender roles and groups of variables on software and identity, and attitudes to Western IT products and services.

As a result of extensions to the Diamond Model, as well as the relative impacts of specific variables, the relevance of segments to Interaction Design for developing countries was given a revised order on the basis of their frequency and strength of influence (numbers in brackets are the initial rankings, as above):

1. Social Structure and Interaction (3)
2. Language and Semiotics (1)
3. Economic Processes (5)
4. Political Processes (4)
5. Material Culture (2)

Two segments were thus raised by two positions in the ranking, and two dropped by one or three places. Much of this was due to the importance of social factors in shaping how users evaluated their experiences, with the economic factors in a developing country having a strong impact on access to IT and associated IT literacy. Material culture and the cultural preferences associated with features such as colour, icons and layout was a minor factor, in contrast to much of the HCI work on culturability.

Figures 2, 3 and 4 show the Diamond Model as extended after the three Jordanian studies. The Diamond Model is presented as a hierarchy of segments that contain objective or subjective groups of variables, and these groups in turn contain territorial variables. A three level numbering system relates variables to their segments and groups. In Figures 2, 3, and 4 a rectangle indicates a cultural or territorial variable that had been identified before in the literature, and an oval indicates a new variable that we did not encounter in our surveyed generic or HCI literature on culture. There are three colours of the ovals; each colour according to the three studies by the authors: white for variables added in first study, dark grey for variables added in the second study, and light grey ovals for variables added in the third study. No Jordanian instances were found for groups or variables in italics. Most of these had their origins in the general literature on culture (e.g., Hofstede 1980). The relevance of these cultural variables to HCI is open to question, as many originate in research that predates the emergence of interaction design. Del Galdo and Nielsen (1996) drew attention to many of these cultural variables, but the relevance of their comprehensive general survey 15 years on appears to be diminishing.

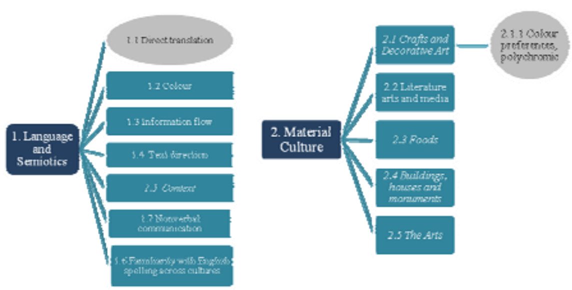


Fig. 2. Language and semiotics and material culture variables in the Diamond Model

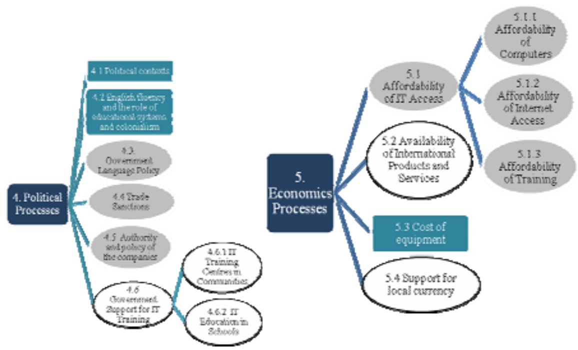


Fig. 3. Political and economic process variables in the Diamond Model

4 Culture, Territory and Technology

In the Middle East, Information and Communication Technologies are spreading very quickly and developing more rapidly than previously. Some usage problems that arise are specific to Jordan, but many probably apply to much of the Middle East and other developing countries too.

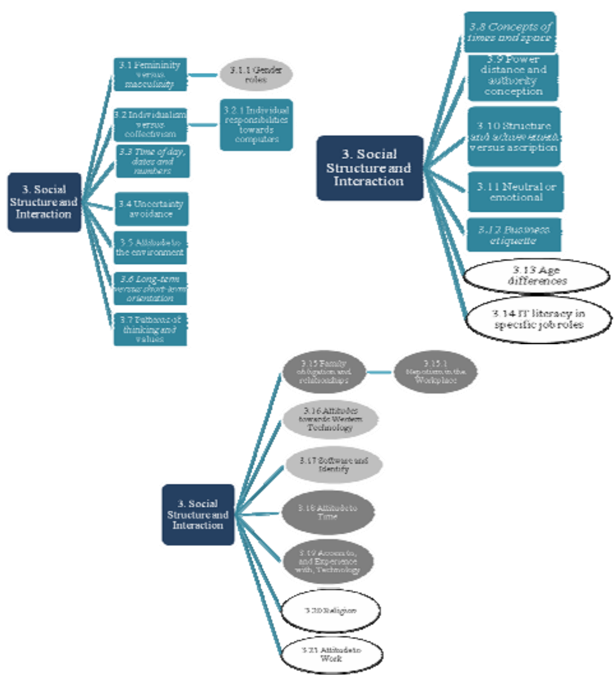


Fig. 4. Social and structure interaction variables in the Diamond Model

Designers of systems for these territories need to understand the broad range of cultural and other territorial factors that shape acceptability. For example, political issues in Saudi Arabia (KSA) and United Arab of Emirates (UAE) have caused problems for Research In Motion's Blackberry mobile products and services, which required some changes in operations. This adds to the evidence that a narrow focus in HCI on cultural factors is inadequate. However, some HCI4D work may be focused too much on economic factors, and not enough on cultural or political variables.

Designers and developers aim to introduce new technologies that can be easy to use and meet users' needs and preferences. As the Diamond Model has expanded and been populated, it has become clear that cultural preferences for user interface features have less impact less frequently than cultural variables that influence users' evaluations of their usage experiences. We should not transfer Western expectations on the relationship between usability and users' affective responses to other cultures. Culture differences result in different responses to usage difficulties, which in turn are due to complex interaction of cultural variables relating to attitudes to IT, attitudes to work, reward systems, attitudes to time, social relationships within the workplace, and further sociocultural factors. Where cultural values are similar between the West and Jordan, as with the impact of students missing deadlines for handing in assessed work, then usage difficulties can have similar adverse impacts. However, our second study revealed that our larger sample of Jordanian students were much more likely to be angry with themselves, rather than angry with an inanimate computer, when compared to Lazar and colleagues' (2006) US sample.

Designers need to understand users deeply by understanding their culture and their motivations and behaviors when using technology. A Diamond Model populated with territorially specific instances provides a basic form of comprehensive information for designers, but we have experimented with *dramatic sketches* as resource for more effective communication of cultural differences to interaction designers. The hope is that a small set of dramatic sketches can cover most instances of cultural variables for a territory, with the remainder covered by briefer micro sketches. We feel that this could be a more effective way of communicating relevant territorial factors to designers than the academic presentation of the Diamond Model in the figures.

5 Conclusions

The main aim of our Diamond Model is to support designers and software developers in the future to understand the difference between users from different cultures and how cultural and other territorial variables can affect not only design preferences, but their usage behaviours and their evaluation of their own user experiences. The Diamond Model is based on a more appropriate physical metaphor than existing meta-models of culture, and extends beyond culture to a relevant range of territorial variables. The model can be readily extended and populated with examples from literature surveys and field studies in relevant countries. The Diamond Model is also a basis for innovative design resources such as dramatic sketches that can combine concrete examples into contextualised narratives to efficiently communicate the impact of cultural differences on the user experience.

References

1. Barber, W., Badre, A.: Culturability: The merging of culture and usability. Presented at the Conference on Human Factors and the Web, AT&T Labs (1998)
2. Chetty, M., Grinter, R.: HCI4D: hci challenges in the global south. In: CHI 2007 Extended Abstracts, pp. 2327–2332. ACM, New York (2007)
3. Del Galdo, E.M., Nielsen, J.: International User Interfaces. Wiley, Chichester (1996)
4. Hofstede, G.: Culture's Consequences International Differences in Work-Related Values. Sage, USA (1980)
5. Kluckhohn, C.: The study of culture. In: Lerner, D., Lasswell, H.D. (eds.) The Policy Sciences. Stanford University Press, Stanford (1951)
6. Lazar, J., Jones, A., Hackley, M., Shneiderman, B.: Severity and impact of computer user frustration: A comparison of student and workplace users. Interacting with Computers 18(2), 187–207 (2006)
7. Qirem, F., Cockton, G., Loutfi, M.: Cultural Differences is Severity and Impact of User Frustration. In: IWIPS 2007, Mexico, pp. 9–18 (2007)
8. Smith, A., Yetim, F.: Global Human Computer Systems: cultural determinates of usability special issue of Interacting with Computers. Whole Issue 16(1) (2004)
9. Triandis, H.C.: The analysis of subjective culture. John Wiley, New York (1972)
10. Trompenaars, F.: Riding the Waves of Culture: Understanding Culture Diversity in Business, London Nicholas Brealey (1993)
11. Victor, D.: International business communication. HarperCollins, New York (1992)

GooGreen: Towards Increasing the Environmental Awareness of Households

Ruud Mattheij[1], Lindsy Szilvasi[1], Lorraine de Beer[1], Kartini Rakiman[1], and Suleman Shahid[2]

[1] Department of Information and Communication Sciences Tilburg University, Tilburg, the Netherlands
R.J.H.Mattheij@uvt.nl, L.M.F.Szilvasi@students.uu.nl, L.M.A.T.deBeer@uvt.nl, K.C.Rakiman@uvt.nl
[2] Tilburg centre for Cognition and Communication, Tilburg University, Tilburg, the Netherlands
s.shahid@uvt.nl

Abstract. In this paper, we present an interaction design exploration into the possibilities of using a computer application to increase the sustainability of households in an enjoyable manner. This is achieved by providing household members with personal information regarding their energy consumption and its related costs. This application also introduces a competition element that links the energy consumption in every household member's room to their personal user account in order to motivate users to compare their energy consumption with other household members. Early results show that the system indeed enhances the environmental awareness and that users are pleased with the idea of competition-based energy savings.

Keywords: Saving energy, sustainability, energy consumption, advice, control, competition element, environmental awareness.

1 Introduction

Sustainability and environmental awareness have been on the policy agenda for decades, resulting in many studies aiming to encourage households to reduce their energy consumption. Many of these studies showed that providing household members with information about ways to reduce their energy consumption tends to result in higher awareness, but does not necessarily change their behavior or actually aids them in saving energy [5]. Likewise, it is concluded that rewards can effectively encourage energy conservation [2], although mostly with short-lived effects [1].

Furthermore, related research indicates that improved feedback on energy consumption might provide households with a tool to control their energy consumption and, ultimately, improve their energy savings. It is also indicated that it is more effective to provide household members with electronic feedback of energy consumption than giving this information in hardcopy [8]. It is concluded that the most successful feedback combines the following features: long and frequently given feedback, the

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 500–509, 2011.

provision of an appliance-specific breakdown, presented clearly and in an appealing way and using computerized and interactive tools [3][4][6]. Based on these design principles, a number of applications have been designed for persuading households to control their energy consumption [7], although most of these applications only incorporate small set of design principles and miss many crucial ones, which play a key role in changing people's attitude towards energy consumption.

Therefore, it is hypothesized that an application that incorporates proven design principles and that allows households to view and control their energy consumption with a single application would provide household members with better insights concerning their energy consumption. It is also hypothesized that the introduction of a challenging competition element in the application might keep the application interesting to users for a longer period of time. It might also help in reducing boredom that rises over time. This would therefore increase the household's sustainability and may also lead to financial savings, thereby making the application financially attractive to households.

In this paper we present the GooGreen application, which aim is to test this hypothesis. The idea is to collect and combine all key design principles that are related to persuasive interfaces and sustainable technologies in one application and assist users in saving energy and maintaining a sustainable lifestyle in an enjoyable and extremely user-friendly manner.

2 User Research

From literature, it was clear that there is a need for an application that could help household members in saving energy. It was also apparent that changing the behavior of people towards energy consumption is a real challenge. This raised the question how household members use or waste energy and how they could be convinced to change their behavior. We therefore performed a proper user research, which consisted of observations and semi-structured interviews. We consulted literature and conducted a brainstorm session to prepare our questionnaire and questions.

These questions were combined into an interview, which was tested during a pilot test. Feedback from this test helped us to improve the interview. All questions in the interview were divided into four different categories/themes: *environmental awareness* (the degree and relevance of sustainability in the household), s*ystem content* (preferred and/or required features of the application), s*ystem layout* (the appearance/attractiveness of the application) and s*ocial activities* (regarding a competition element of the application).

After a pre-pilot test, preliminary user studies were conducted. These user studies consisted of semi-structured interviews with 12 participants in their own familiar (home) environments: 7 men (average age 39) and 5 women (average age 41) using the interview and topics mentioned above.

From our user studies, we were not only able to confirm the previous findings [2][3][4], but we also found interesting information about Dutch households. A few of the important findings are as follows:

- Participants are willing to increase their environmental awareness and sustainability, also if this would require them to use a computer application.
- They find it morally important to live sustainable, even if it requires changing some of their habits.
- Using such an application should not bring any major discomfort and should not be an extra overhead.
- This application should work as a supporting tool for enhancing environmental awareness.
- They should be able to use this application using a single device, at any time and at any place in the house, for controlling all appliances.
- Participants would like to be environmental friendly, although the most important reason for participants to use such an application will be to reduce costs and have a complete financial overview for the year.
- The real return on investment will be less expensive yearly bills. Many participants are worried about the increasing costs of energy and are annoyed by the extra amount of money they have to pay at the end of the year for extra energy consumption.
- A clear overview with proper feedback is extremely important for controlling the energy usage.
- Concerning the social activities, the main finding was that the majority of the participants would like to have a competition element in the application.
- The participants mentioned that it would be much more fun when they can compete with children and that children will find a fun element in this feature. They will also start contributing to a more sustainable lifestyle.

3 Design Ideas

The results of the user studies were discussed during a brainstorm session. The main results from this brainstorm session were that we had to narrow the scope of energy consumption to gas and electricity consumption. Likewise, the results indicated that we had to make the application very personal. Users should be provided with advice about their energy consumption, while they should also be able to control it. A competition element should be used in order to stimulate users to compare their energy consumption.

4 Final Design

The design ideas led to the development of a preliminary design of a paper-based low-fidelity prototype. The paper prototype was tested with 5 participants, which gave us interesting results about the contents and the visualization of the application. After the low-fi prototype, based on users' feedback, we decided that the application should contain clear schemes, symbols and icons, rather than text.

Likewise, we decided that the application should be touch-sensitive and that it should be possible to run it on a tablet pc. It was also decided that different layers of the house should be controllable; the entire house, a single floor or just a single room. The competition element was implemented by creating cartoon-like trees, which are connected to unique user accounts in the application.

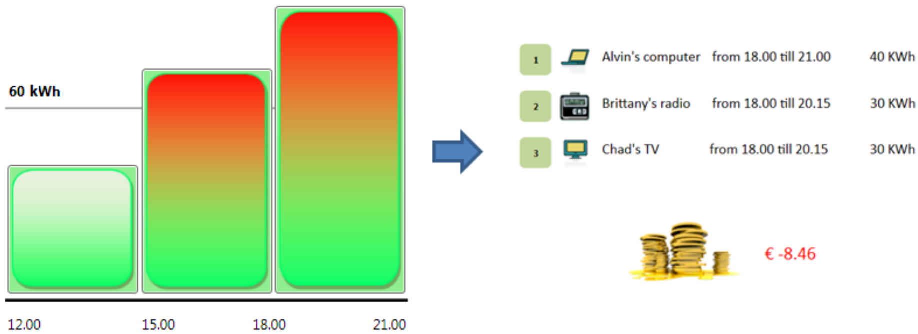


Fig. 1. The energy consumption on a particular period of time and the corresponding energy consuming devices

The final design contained three main function blocks: the *Advice* function, the *View & Control* function and the *Personalize* function.

The Advice function informs users about the amount of energy that has been consumed by devices in the user's household during a certain period of time, using m^3 for gas and kWh for electricity. "Gas" or "electricity" can be selected to consult all corresponding devices. The default backtrack period is 24 hours, but any length of time can be selected. Energy consumption is presented as a bar chart graph. A toolbar is available to select different periods of time, e.g. the last 24 hours or the last week.

More specific periods of time (e.g. from 12.00 to 15.00) can be consulted by selecting the corresponding bar in the bar chart graph to see which devices were powered on during this period of time. Finally, GooGreen calculates how much money the energy consumption has cost during this period of time. This informs users about devices that could be powered off more frequently in order to save energy. See fig. 1.

The View & Control function informs users about the current power status (powered on or off) of devices. Devices can actually be powered on or off by this function. There are three different levels of control (see fig. 2), which are comparable with zooming in on a house:

Entire house; this is the top level of three recursive levels. GooGreen shows a simplified view of the house, displaying the number of gas and electricity consuming devices that are powered on at that moment on every floor of the house. This level of control should allow users to quickly power off all gas and/or electricity consuming devices in the house.

Single floor; on a floor level, this level of control is analog to the entire house level. There is a slight variation in the display of these levels, as pictured in fig. 2.

Single room; this level shows a list with devices and their power status (powered on or off) in that particular room. This level allows users to power specific devices on or off. Users select the gas icon in a room on the floor-level to navigate to a menu with an overview of all gas consuming devices in that particular room.

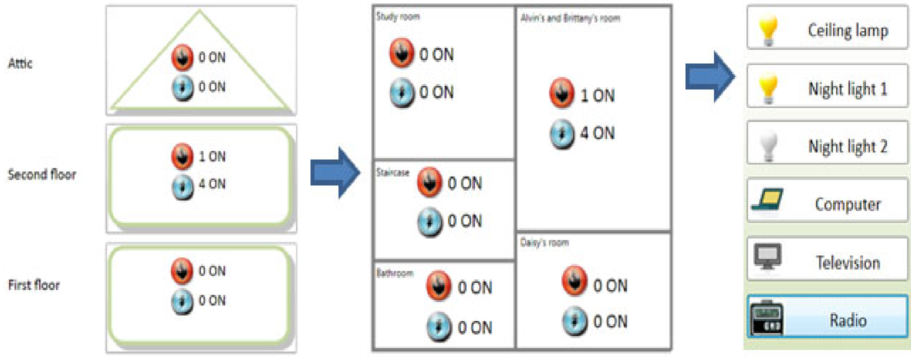


Fig. 2. The three level flow: house level (left), floor level (middle), room level (right)

In this overview, users can power these devices on or off. Likewise, selecting the electricity icon in a room on the floor-level navigates users to a menu with an overview of all electricity consuming devices in that particular room.

The Personalize function allows users to configure GooGreen according to their wishes. It is divided into two different kinds of settings: application settings (change GooGreen's appearance) and household settings (add, remove or change accounts). A virtual keyboard is available for text input. The GooGreen prototype uses an aura effect for highlighting interactive objects (e.g. a button) when users hover above them with their finger(s).

Every household member that uses GooGreen has a personal account. Every account has its own energy tree, a visual representation of the amount of energy that is consumed in the room of the account's owner. Higher energy consumption results in a less vital and unhealthy looking energy tree. Likewise, less energy consumption results in a more vital and healthy looking energy tree. See fig. 3.

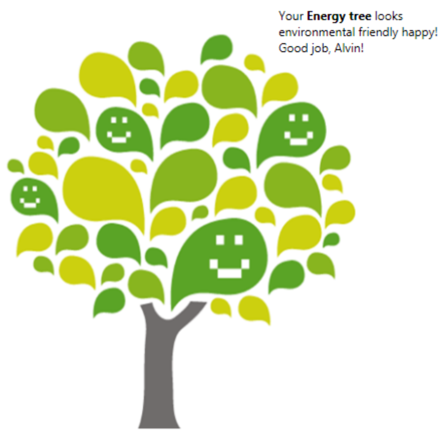


Fig. 3. The energy-tree provides an indication of the amount of energy a household member has used during a certain period of time. If a household member consumes more energy, the tree will look less healthy.

The healthiness of the users' energy trees can be mutually compared to find out which user consumed least or most energy. See fig. 4.

5 Implementation

A high-fidelity prototype of GooGreen was developed using Visual Studio 2008. GooGreen uses the Windows Presentation Foundation (WPF) for its frontend and C# for its backend. The high-fidelity prototype runs on Microsoft Windows tablet PCs. It requires the .NET 3.5 framework; a free software framework from Windows that enabled us to create the different graphic effects of GooGreen.

The information shown in GooGreen screens does not represent an actual house; the household, its members and its energy consumption are entirely fictional in order to prevent privacy related issues. This application requires devices that can be powered on or off remotely and that are easily accessible. These devices should be connected to the gas or electricity consuming household appliances that should be controlled by GooGreen.

6 Evaluation

The evaluation consisted of the heuristics evaluation and the usability tests. For the heuristics evaluation, we used the ten heuristic principles of Nielsen. Three expert reviewers tested the application for 10-15 minutes each. The usability test provided an indication of the quality of our solution by testing the time the participants needed to complete a task and the way they thought about GooGreen, thereby providing both quantitative and qualitative data. In the first phase of the usability evaluation, we wanted to measure the usability and acceptance of GooGreen. In the second phase, we would like to see its long-term impact on users' behaviors. The first evaluation consisted of a heuristics evaluation and a usability test.

Six people, 4 men (avg. age 31) and 2 women (avg. age 27) participated in our study. The test consisted of a pre and post-questionnaire and a task book with 11 key tasks.

Fig. 4. The energy-trees can be mutually compared with other household members, thereby creating a competitive element. This specific example shows the energy trees of four different users with different states of 'environmental friendliness'.

7 Results and Discussion

The mood test measured the participants' mood prior to the test by using a 1 to 7 Likert-scale; 1 means 'feeling very bad' and 7 means 'feeling very good'. The pre-questionnaire was used to ask general questions concerning the users' overall experience with computers and their affinity with sustainability. The results indicated that participants had a reasonable affinity with computers, used computers for work, study and leisure time and that all participants were positively interested in sustainability.

The assignment consisted of 11 tasks. The time necessary for each task was measured and then compared with baseline time, which was set by the help of experts during the heuristic evaluation. The time necessary for each task was measured and compared with baseline time. Overall, all participants were able to finish their tasks in a reasonable amount of time. There were no unnecessary delays. Expertise in computer usage had a positive impact on completing the tasks, although this was not significant.

The first task consisted of starting up GooGreen, after which the participants were asked to retrieve information about the period in which they consumed most energy. All participants indicated that these tasks were easy to perform; it took most participants less than a minute to understand the advice function. After retrieving this information, all participants were asked to actually influence the energy consumption by performing a variety of scenarios, based on switching certain electric and/or gas consuming devices on or off. It took most participants about two minutes to understand this control function, which was mainly caused by the number of levels in the house (entire house, entire floor and entire room) the participants had to search in order to perform their tasks. Personalizing GooGreen to the preferences of the individual participants was a minor problem for most users. It took some participants a few minutes longer than the baseline to add a new user to the application. Consulting the high scores screen in order to compare the energy tree of the participants to other household members' energy trees was very convenient; all participants performed this task without problems.

After these experiments, a post-questionnaire was used to measure the attitude of participants towards GooGreen by asking users to what extent they agreed with the given statements. Their level of agreement was again measured on a 1 to 7 Likert-scale; 1 means 'totally disagree' and 7 means 'totally agree'. Table 1 indicates that some tasks are apparently missing in the prototype. The participants' feedback was that they would appreciate "water consumption" to be included. Secondly, there was also a strong demand to include a feature to power on devices at a pre-programmed time.

It is also clear that the ratings for the first two questions are relatively low. This is due to the fact that the application was not fully functional as it was not directly attached to the physical devices. This had a little impact on the user experience and evaluation. Although users did not show great willingness to buy the product immediately, they totally understood the concept and most of them realized that GooGreen's interface is the main thing.

Table 1. The statements that were served during the post-questionnaire and the corresponding level of agreement of the users

Statement	Average score
Now that I have tested GooGreen, I would purchase it myself.	4.2
Now that I have tested the application, I would recommend it to others.	4.8
The application looks well-organized.	6
The application looks appealing.	5.8
The application is easy to use.	5.6
I can perform all tasks I would like to perform with this application.	1.6
I think it is a good idea to implement the energy tree in the application as a competition element.	6
The energy tree is a fun part of the application	6.2
The energy tree looks appealing	6.2

When compared to the mood test prior to the experiments, the results indicated that most users felt more satisfied and happier after using the GooGreen. They also thought that it is morally good to save energy and this feeling had a positive impact on their mood. According to the average scores, the application appeared to be appealing, well organized, easy to control and easy to use. Participants did not feel the application to be confusing. The experience with the energy tree was appreciated because of its appealing looks and the fun part of its competition element. The competition element turned out to be a very important feature and all participants unanimously appreciated it. It improved the application and was a motivating feature. Most participants considered it to be an effective way to stimulate the long-term use of the application. We learned from the user tests that this element should not be present at an individual level, but at a house level, community level or town level and that it may be spread among friends, especially in the case of children. They thought that this element is really motivating. On the other hand, they also showed a strong interest in privacy and the security of the system. They really wanted to have a controlled application, in which they decide ‘who can see what’.

The results confirm the hypothesis that an application that allows households to control their energy consumption from one central point at home would provide household members with better insights concerning their energy consumption. Our project seems to support Fischer’s conclusion as stated in the introduction.

Because of time constraints, the long term effects of the use of GooGreen could not be studied, although it is felt that this is an essential topic for a complete assessment of GooGreen. The data from the user tests is limited because the usability tests were performed with only few participants. Therefore, it is possible that results from usability tests with larger groups differ from our results. In any case, the results, especially the combination of the heuristics evaluation and usability test, do show a strong trend that not only confirm existing findings, but also give new insights about energy consumption and saving.

GooGreen differs from other attempts to increase the sustainability of households because it contains a competition element (the energy tree concept and account system) and it allows users to influence energy consumption in their household in a systematic manner with minimal interaction. This solution also offers clear information about the costs of the energy consumption and it can work as an additional tool to manage its financial aspects. This approach makes consuming and saving energy very personal. It challenges users to become more involved with the sustainability of energy and resources.

This study also revealed that changing behavior is a long-term process and that this can only be achieved by a number of short-term milestones and rewards. Financial benefit is a major reward and an immediate feedback, which continuously updates users about their financial savings. Therefore, this will be the first real step towards a sustainable lifestyle.

8 Conclusion and Future Work

In this paper, we presented an interaction design exploration into the possibilities of using a computer application to increase the sustainability of households in an enjoyable manner. For this purpose, we developed the GooGreen application, which was designed in an iterative manner. Users did not only participate in the design of the application, but also evaluated its working version. Participants stated that GooGreen's presentation of information and advice regarding energy consumption increased their awareness about energy consumption considerably. Participants liked the idea to view and control their energy consumption in the household with a single application. The participants acknowledged the value of the competition element (the energy tree concept and account system) as a solution to involve more people in saving energy in a household. The initial use of this application was appreciated by most users.

This application, however, is still limited in a number of ways as it was started as a student project during a course. Therefore, the time period for development and evaluation was a little short. We could not build a fully functional system, which actually connects with physical devices. Likewise, we did not allow users to try this application in their daily routine on long-term basis.

For future research, we would like to extend this application's development and evaluation in a number of ways. First of all, we recommend a longitudinal investigation to determine the attractiveness of this solution. We also want to perform a cost-benefit analysis to study the relationship between buying this application and the amount of money that can be saved by the help this application. Another important improvement is to develop an interface between this application and physical devices.

References

1. Abrahamse, W., Steg, L., Vlek, C., Rothengatter: A review of intervention studies aimed at Household energy conservation. Journal of Environmental Psychology 25(3), 273–291 (2005)
2. Bang, M., Torstensson, C., Katzeff, C.: The PowerHhouse: A Persuasive Computer Game Designed to Raise Awareness of Domestic Energy Consumption. In: IJsselsteijn, W.A., de Kort, Y.A.W., Midden, C., Eggen, B., van den Hoven, E. (eds.) PERSUASIVE 2006. LNCS, vol. 3962, pp. 123–132. Springer, Heidelberg (2006)

3. Fischer, C.: Feedback on household electricity consumption: a tool for saving energy? Energy Efficiency 1(1), 79–104 (2008)
4. Fogg, B.J.: Persuasive technology: using computers to change what we think and do. Ubiquity 12, 89–120 (2002)
5. Heijs, W.J.M.: Household energy consumption: Habitual behavior and technology, pp. 149–157 (2006)
6. IJsselsteijn, W.A., de Kort, Y.A.W., Midden, C., Eggen, B., van den Hoven, E.: Persuasive Technology for Human Well-Being: Setting the Scene. In: IJsselsteijn, W.A., de Kort, Y.A.W., Midden, C., Eggen, B., van den Hoven, E. (eds.) PERSUASIVE 2006. LNCS, vol. 3962, pp. 1–5. Springer, Heidelberg (2006)
7. Kim, T., Hong, H., Magerko, B.: Designing for Persuasion: Toward Ambient Eco-Visualization for Awareness. In: Ploug, T., Hasle, P., Oinas-Kukkonen, H. (eds.) PERSUASIVE 2010. LNCS, vol. 6137, pp. 106–116. Springer, Heidelberg (2010)
8. Wood, G., Newborough, M.: Dynamic energy-consumption indicators for domestic appliances: environment, behavior and design. Energy and Buildings 35(8), 821–841 (2003)

User Experience of Social Bookmarking Tools

Enric Mor[1], Nuria Ferran[2], Muriel Garreta-Domingo[3], and Juan-Antonio Mangas[3]

[1] Computer Science, Multimedia and Telecommunication Studies -
Universitat Oberta de Catalunya, Rambla del Poblenou 156, 08018, Barcelona, Spain
[2] Information and Communication Sciences Studies - Universitat Oberta de Catalunya,
Rambla del Poblenou 156, 08018, Barcelona, Spain
[3] Office of Learningn Technologies - Universitat Oberta de Catalunya,
Av. Tibidabo 39, 08035, Barcelona, Spain
{emor,nferranf,murielgd,jmangas}@uoc.edu

Abstract. Information and knowledge society brings a new context where technology enhanced tools are key elements for being able to find, evaluate, use and communicate information effectively and efficiently [1]. Bookmarking tools could be the essential tools for supporting information behaviour, specifically information managing and communication. This paper analyses the user experience of existing bookmarking and social bookmarking tools in an e-learning environment. The educational setting provides the required environment to truly study these tools, since their success is not only in the ease of storing, tagging and sharing resources at a given moment in time but in how these resources will be retrieved when needed in the future. In this paper we present a functional analysis and the usability inspection of the tools that support the management and usage of information resources both during short and long terms.

Keywords: Information society, information skills, information management, bookmarking tools, usability evaluation, e-learning.

1 Introduction

In a networked society, most people have the need to search, find and store information over the Internet. Users browse the web in search of content and services of their interest. Search engines like Google or directories like Yahoo! help them find what they are looking for. Using these services people locate interesting resources and access them, but when surfing the web, people also have to deal with things such as URL names, addresses, links and other things more related with the organization and infrastructure of the web than with their needs as users. All these issues are not easy to remember and, sometimes, make the browsing experience stressing.

Social networks and tools like Facebook or Twitter provide people with a way to interact with each other and share comments, images and experiences. In addition to that, these tools are an easy way to share links to web pages, articles, videos or other types of resources and, at the end of the day, each user has a great amount of references of resources to look, watch, read or listen. It is not an easy thing for the users to deal with all these resources, remember the ones they find interesting or store the URL they want to read or access later.

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 510–516, 2011.

Some years ago, bookmarking tools promised a solution to deal with this situation and presented themselves as key tools for users to achieve their information needs. The success and experience provided by these tools should be evaluated taking into account the long term period, that is, not only the ease of use of bookmarking resources but how these can be retrieved when needed again.

In the education area, information needs are even more important. Internet resources are part of the contents that students and teachers use to learn and teach. There are subjects – such as those related to or affected by information technology - that require a continuous update of their learning contents. As a result, the use of a closed format for these learning materials is not optimal in these type of courses, since in a very short period of time the material would be outdated. The information and knowledge society has introduced technology enhanced learning tools that students have to learn and, therefore, require to develop information competencies in order to do so. On the other hand, teachers also face new challenges and they must change the role from knowledge dispenser to knowledge facilitator. This change in the educational model is also changing the channels for transmitting knowledge, mainly textbooks. All these aspects promote a direct use of resources available in the real world, especially on the Internet. This is intended to encourage students to select the networked resources directly, use them, reuse them and share them with the rest of the educational and learning community. Given this new educational model in which the exchange of content is encouraged, some challenges and shortcomings must be solved.

In this educational context, not only technology enhanced learning tools are used but also existing Internet tools that can be used for teaching and learning. Good examples of this are the bookmarking and social bookmarking tools, like Delicious, Evernote or Diigo. These kind of tools are very good to tag and store links and even to share the links and build common folksonomies based on used tags. Nevertheless, most of the times, the experience to retrieve stored resources is not as good. From an information competencies point of view, the discovery, evaluation, storage and retrieval of resources are the key elements. Users (students) are becoming good in discovering and evaluating Internet resources. Tools are good and provide a good experience to store and use such resources but they should provide better ways to retrieve and use the resources later on.

The use of bookmarking tools and the features they offer is an essential element in the paradigm shift in education. They offer a set of advantages and drawbacks not only intrinsically because of their design and usability but also in relation to people information skills. Following sections present a deeper insight on these aspects. Section 2 presents the necessary information skills in the information society and their relation with bookmarking tools, Section 3 presents the study and evaluation of bookmarking tools. Section 4 shows the current situation regarding information needs and uses, and finally, Section 5 presents the conclusions of this work.

2 Information Needs and Bookmarking Tools

Information society is a networked society where the development mode is having a rapid process of change. It is different because the generation, process, and transmission of information have become sources of productivity and power of this society [2].

In order to educate qualified people for the information society, it is essential to rethink teaching and learning, make it student-centered, thus building upon the acquisition and development of skills rather than content acquisition [3]. Information literacy is seen as essential for evolving the dominant paradigm of preprocessed information to a new learning paradigm in which students take over the capabilities to self-manage their learning throughout life. In addition, textbooks, lectures, exercises and other resources should aim to rely the learning process on resources available in real life and real work situations [4].

Therefore individuals with information skills should be able to [5]: classify, store and manipulate the generated or retrieved information retrieved and, recognize information literacy as a prerequisite for learning throughout life. More concretely, students with information skills should be able to 1) determine the nature of their information needs, 2) access to information needs effectively and efficiently, 3) evaluate information and its sources critically, 4) incorporate selected information into their knowledge base, 5) use information effectively to achieve the aim, 6) understand the economic framework, legal and social information usage, access and use of legal and ethical manner.

It is important for people to have tools to deal with information needs and how to find information, but, what is more, tools for managing information and evaluation. Bookmarking tools are specific tools for reference and link management. These tools can be part of the solution for managing and evaluation. Bookmarking and social bookmarking tools are appropriate applications to reference and share the knowledge available on the net. These tools are mainly characterized by: 1) save and retrieve links to online resources, 2) tag and comment on resources, 3) rating of resources, 4) online storage. Normally there is an associated storage service and all information is stored in the cloud and, thus, a user can access the same information, at any time, from different computers or devices and further information is safe from possible malfunction of the computer.

In addition to that, bookmarking tools can provide a layer for social interaction, thus encouraging exchange of resources and knowledge and the interaction and with others. The social layer is very useful in different aspects: sharing of resources with people who have similar needs and interests, knowledge about the "popularity" of a resource and construction of folksonomies.

Currently there are many bookmarking tools, most of them online through a web interface and free of charge. Some of the most interesting tools are:

Table 1. Main bookmarking and social bookmarking tools

Tool	URL
Delicious	www.delicious.com
Blogmarks	www.blogmarks.net
Blinklist	www.blinklist.com
Mister Wong	www.mister-wong.com
Evernote	www.evernote.com
Furl	www.furl.net
Diigo	www.diigo.com
Google Bookmarks	www.google.com/bookmarks/
Xmarks	www.xmarks.com
Historious	www.historio.us
Evri	www.evri.com

It is also interesting to note that there are tools that were originally created as bibliographic references managers and now also provide link and URL management. Some of these tools are: CiteULike (www.citeulike.org), Connotea (www.connotea.org), BibSonomy (www.bibsonomy.org) or Zotero (www.zotero.org).

3 Tool Evaluation

The current situation in which more educational content is in a digital format and distributed through the net, leads to the need of useful and usable tools to manage electronic resources. The management of these resources is mainly based on saving the web address of the resource and, on the other hand, retrieving resources previously stored. The user experience of the bookmarking tools is an important part of their success and, for this reason; their usability evaluation is of great importance. In this work, the main goal of the evaluation of bookmarking tools was to detect their strengths and weaknesses, both in terms of ease of use and utility, and thus diagnose their best use in an educational context.

The evaluation of tools in this work was carried out using a heuristic evaluation, a classic method for usability evaluation. This method was chosen because it is a reliable method for discovering usability issues and provides information about the shortcomings and strengths of the tools studied. Heuristic evaluation is an usability inspection method that is used to identify usability problems of interfaces [6]. This type of evaluation is usually carried out by one or more experts. The evaluators study the interface and confront it with a set of heuristics or rules previously established. Heuristic evaluation was defined by Nielsen and Molich in 1990 to study and analyze the compliance of an interface with usability principles. The heuristic principles are usually structured in the form of checklists, thereby facilitating the evaluation. For each heuristic, indicators are formulated as questions, where an affirmative answer indicates the absence of a usability problem.

To evaluate the bookmarking tools we involved three expert evaluators. They decided to use the heuristics defined by J. Nielsen [7] as they cover the main aspects to evaluate and they provide sufficient flexibility to adapt and define the indicators of each principle to evaluate the characteristics of the tools in educational contexts. The evaluators also selected the tools to analyze and each expert assessed all the selected tools. The indicators for each principle were measured in terms of 0 = never meets the requirement, 1 = sometimes, 2 = meets the requirement.

From the available bookmarking tools, experts selected four tools: Delicious, Google Bookmarks, Evernote and Diigo. The main features of bookmarking tools are storing and retrieving found links to Internet resources. The selected tools provide these basic options, and also other features that are considered important, such as: remote storage of information, sharing resources with other users, post links to resources and the capacity to import and export the stored data. With these capabilities, these tools are suitable for educational contexts in general.

4 Evaluation Findings and Discussion

In the visibility of system status principle, all tools got similar results. Evernote gets a higher score because provides more feedback to the user. From the visibility point of view the four tools are adequate for the management of information in educational contexts. In the second principle, match between system and the real world, all tools get good score in the evaluation. Experts highlight Diigo feature about annotations on web pages being a clear metaphor of the real world. The interface and interaction design show an effort to make things intuitive and provide a good experience to the final user. This is an important aspect in educational settings because the user's goal is to read and learn the bookmarks and not the management of the URL. The less time devoted to learning and using the tool the more time for learning they will have. Related to the user control and freedom principle, all the analyzed tools provide effective solutions to go back and to return to previous state. The four tools are rated similarly but Evernote got a better result thanks to the software options it provides both for browsers, mobile devices and desktop operating systems. This software enhances the bookmarking tool and provides great versatility in its use and this is a key aspect of tools in an educational context. Students and teachers can use the tool according to their requirements, device of preference and features needed. In the consistency and standards heuristic, Delicious and Diigo obtained a highest score because its visual design and the carefully elaboration of details. Consistency and standards reduce cognitive load in the interaction with systems and this is an important aspect in educational contexts where the intellectual effort should be devoted to the topic or subject of learning itself. Related to error prevention, all evaluated tools receive the same score. Each one is simple enough not to allow the user to carry out errors and, at the same time, offer complete and sometimes complex features and options. Nevertheless, users cannot be completely controlled and always can carry out unwanted actions. The results obtained in the evaluation of the recognition rather than recall principle place Google Bookmarks in an inferior position to the other tools. This is because the tool makes primary use of a textual design against a design that uses images and icons. The main drawback is that Google Bookmarks does not make clearly visible the major options needed in a bookmarking tool. The heuristic principle about flexibility and efficiency of use is related with the ease of use of the tool. The tools need to be useful and easy to use for novel users as well as for experienced users. In this evaluation, Evernote has the highest punctuation as it offers a set of functionalities richer than the rest of the tools. Both Evernote, Delicious and Diigo are tools that offer a great flexibility. The principle about aesthetic and minimalist design means that bookmarking tools should have the needed information and elements not essential produce interferences in the user. Evernote tool has the lowest punctuation as its interface has a lot of elements, but this can also be as it is the tool with more functionalities. Google Bookmarks gets very good results on this principle but offers fewer options and features. In the principle about help users recognize, diagnose, and recover from errors, all tools show errors if occur and provide explanations for users to recover from this situation. The design of the tools is made in a way that users choose options that usually do not lead to mistakes. Anyway, the most common errors are related with the erasing of data without intention. Google Bookmarks do offers the possibility of recovering immediately and Evernote keeps in the trash the deleted elements offering the possibility to save the needed data. The help and documentation principle evaluates how easy can be finding information about the software functionalities when needed, if the user can search among the help information and how well structured and organized is the training information. Evernote, is the tool with highest score in this evaluation aspect mainly thanks to the excellent help and documentation section it offers. Delicious also provides a good help section but it is not that well organized and clear. The help and training offered by these kind of tools is very important in educational contexts where additional support is very important for students.

The main finding of the evaluation was that although there are tools better scored than others, all analyzed tools are suitable for use in educational contexts and, nevertheless all of them have advantages and disadvantages. It is very easy to add resources and links on all the tools and almost all of them provide additional software to do so. On the other hand, the recovery of stored resources is a complex task in all tools. Search interfaces are good but users must remember the keywords of the resource and, sometimes, these are very generic. Navigation interfaces are also good and well designed, but the tools provide little support for browsing and finding resources because navigation depends mainly on the classification and labels the user defined when he / she saved them and, therefore, rely almost solely on the methodology used by the users, that is, in their information skills.

The top rated tools are Evernote and Delicious. These tools have interesting features when used to store and manage learning resources. Evernote provides additional functionality and allows to store files, both text and multimedia, and also easily retrieve information from these files. This is an important aspect in an educational setting where not only it is necessary to manage references and links to resources but also to manage content itself. However this should not be the only aspect to consider when selecting a tool. Delicious, in turn, offers two important comparative advantages: its large number of users and the API that allows building a specific software application to access the stored information. Thanks to its number of users, Delicious builds a folksonomy based on the tags and keywords that users have used to label resources. At the same time, other users can take advantage of that when storing and accessing their own resources because the tool suggests the most appropriate labels for each link. Diigo offers to annotate and comment on web pages and resources, a useful functionality in education. Students and teachers can annotate, comment, edit and discuss content on web pages without having to modify or edit the content itself.

5 Conclusions

One important aspect of the bookmarking tools is that although the basic tasks to perform (save, retrieve and search links) are seemingly simple, they hide some complexity. With the current available tools is very easy to store information and, at the same time, it can be a very difficult to find and retrieve this stored information. Consequently, it can be stated that the design of current bookmarking tools favors information behaviors more focused on saving resources for later than using them right later on. As a main consequence, users need to acquire information competences in order to manage in our current society and overcome the shortcomings of available tools.

In an educational context, the user information goals and tasks have a different dimension that may affect the design and usage of these tools. In these settings, students and teachers have specific goals and tasks which are different from other kind of users. Moreover, the social aspects of social bookmarking tools arise as a very good element that allows interaction among students and teachers.

Therefore, the inclusion of social bookmarking tools in these settings require further research and elements to consider such as the specific learning and teaching needs of users – as it has been mentioned –, the characteristics of the learning virtual environment; and the information skills of users related to maintaining the ease of storing resources, prioritizing and promoting their later retrieval using search tools or navigating through taxonomies or folksonomies.

To pursue the initial research presented in this paper, we plan to 1) study of information management behavior of our students and 2) run an initial pilot with a bookmarking tool integrated in our learning environment in order to evaluate its usage. This analysis of the user experience of social bookmarking tools in a specific environment such as e-learning will also help advance in the design of these tools.

Acknowledgements. This work has been partially supported by a Universitat Oberta de Catalunya grant under the educational innovation project RECRE / APLICA2010.

References

1. American Library Association: A Progress Report on Information Literacy: An Update on the American Library Association Presidential Committee on Information Literacy: Final Report. American Library Association, Chicago (1998)
2. Castells, M.: The Information Age: Economy, Society and Culture, vol. 1. The Rise of the Network Society, Blackwell Publishers, Oxford (1996)
3. Geser, G.: Open educational practices and resources - OLCOS Roadmap 2012. Open eLearning Content Observatory Services (2007)
4. Bruce, C.S.: Information literacy as a catalyst for educational change: a background paper. White Paper prepared for UNESCO, the U.S. National Commission on Libraries and Information Science, and the National Forum on Information Literacy, for use at the Information Literacy Meeting of Experts. Prague (2002)
5. SCONUL: Information Support for eLearning: principals and practices (2003), http://www.sconul.ac.uk/publications/pubs/info_support_elearning.pdf
6. Nielsen, J., Molich, R.: Heuristic evaluation of user interfaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Empowering People, pp. 249–256. ACM Press, New York (1990)
7. Nielsen, J.: Usability Engineering. Morgan Kaufmann, San Francisco (1993)

Part VI
Emotions in HCI

ShoeBox: A Natural Way of Organizing Pictures According to User's Affinities

Bojan Blažica[1], Daniel Vladušič[1], and Dunja Mladenić[2]

[1] Xlab d.o.o., Pot za Brdom 100,
1000 Ljubljana, Slovenia
[2] Artificial Intelligence laboratory, Jožef Stefan Institute, Jamova 39,
1000 Ljubljana, Slovenia
{bojan.blazica,daniel.vladusic}@xlab.si
dunja.mladenic@ijs.si

Abstract. In this paper we tackle the problem of personalizing the experience of browsing through digital pictures. We address two questions: how to capture the user's personal affinity for a particular picture and how to visualize a large collection of pictures. We propose a novel approach towards organizing pictures called ShoeBox that aims for automatic capturing of the user's affinity for individual pictures.

Keywords: personal information management, digital photo collection, visualization, ranking, user-centered design.

1 Introduction

Back in the times of analogue photography some of us used to store photos in shoe boxes. When viewing them, we would gather around a table and turn the shoe box upside down, letting the photos scatter on the table. Nowadays, we know from personal experience, that a few weeks trip can result in a photo collection as large as a few thousand photos. Although each picture represents a certain memory like a place we visited or a person we met, not all of them bear the same importance to us.

Since the advent of digital photography many approaches have been taken in organizing large photo collections. Some of them use tagging [1], [2] or collaborative tagging [3], annotating using ontologies [4], automatic feature extraction [5], clustering according to the pictures' metadata like timestamp [6], [7], [8] and various combination of these.

Besides organization, visualization is also an important factor in personal information management. In contrast to the common visualization of pictures in a grid sorted by date or filename, Photofinder [9] sorts pictures according to a similarity measure and various demo multitouch application visualize pictures scattered on the desktop.

In our opinion the existing systems lack the ability to automatically capture the user's emotional aspect and taking it into account when sorting and visualizing pictures. Figure 1 presents a schematic representation of our solution to this problem. During the first time the collection of pictures is viewed the system builds a model of the users'

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 519–524, 2011.

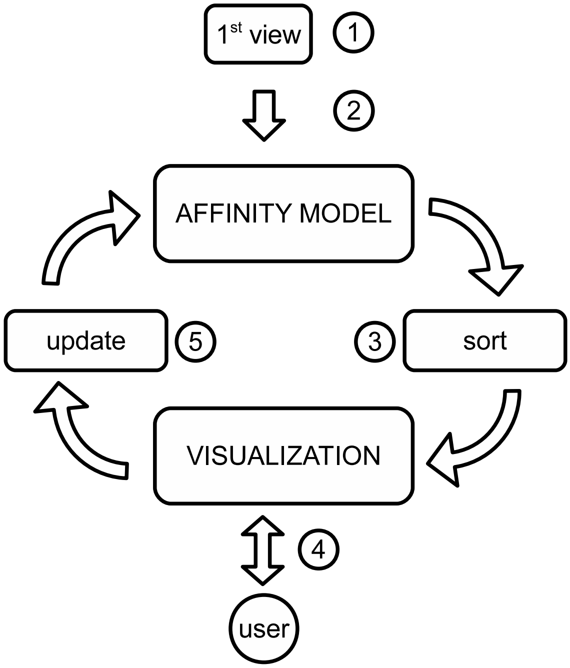


Fig. 1. A schematic overview of ShoeBox; the first time a collection is viewed (1) pictures are sorted according to the pictures' time-stamp. While the user is viewing the collection, the system builds the affinity model based on the time the user spends viewing each picture (2). At each next visualization of the collection, the affinity model is used to sort (3) pictures and visualize them accordingly. During interaction (4) the user's affinities are captured again and used to update the affinity model (5).

affinities. Each time the user views the collection, this model is used to organize it and at the same time the affinity model gets updated based on the user's interaction with the picture. In this paper we focus on using viewing time as a measure for affinity in the affinity model and on a natural way of visualization of large collections of photos.

2 Affinity Model

In defining a measure of the user's affinity for a picture we took a similar approach to the one used in web analytics: time spent viewing a picture. The first time a user views a collection, the pictures are sorted according to the time they were taken on. The user then assigns focus to a picture he wants to view by selecting it. We interpret this as a declaration of affinity for that particular picture and the time spent viewing it (viewing time) is considered a measure of strength of the affinity for that picture. When the picture loses focus, the time elapsed is stored in exchangeable image file format (EXIF). After the first view of the collection the system is already able to sort pictures according to viewing times. The next time the picture is viewed its viewing time will be added. The maximum amount of viewing time that can be added per viewing is set to 30 seconds in case the picture remained in focus while the user went away from the computer.

$$PAS = w_{vt}f(vt) + w_if(i). \quad (1)$$

The introduction of the Picture Affinity Score PAS, described by equation 1, allows for a broader definition of affinity. The affinity score takes into account viewing time and also interaction with a picture (e.g. zooming, rotating) and is calculated as a weighted sum of functions of viewing time (vt) and interaction (i). Besides the affinity score, the algorithm also stores a unique collection ID and the time of the first view into EXIF format. This data is needed when visualizing the collection and updating the affinity model.

Algorithm 1 shows the implementation of decay. This feature foresees the use of ShoeBox for viewing multiple collections simultaneously and is intended to prevent the algorithm to be biased towards pictures from older collections.

To prevent relevant pictures from being completely forgotten, after sorting, an amount of pictures with short viewing times is placed among those with high viewing times. This is done by randomly picking 10% of the 20% less relevant pictures and randomly placing them among the 20% most relevant pictures. We argue that this not only helps the system not to misplace relevant pictures but also aids the user's recall of the memories represented by all the pictures, even the ones he doesn't like.

Algorithm 1. Visualization in ShoeBox with the application of decay based on current time (*CT*) and decay speed (*DS*)

```
if picture doesn’t have EXIF tag 1st_view_time then
  generate collection_ID
  sort according time-stamp
  for picture in collection do
    picture.1st_view_time • CT
    picture.collection_ID • collection_ID
  end for
else
  for picture in collection do
  apply decay:
  AS=picture.AS exp(-DS (CT – picture.1st_view_time))
 picture.AS • AS
 end for
 sort according AS
 apply recall aid
end if
calculate center points
apply random rotation
```

3 Visualization

As mentioned before, we consider scattering pictures on a table to be a natural way of browsing through a large collection of photos. We propose something similar for viewing digital pictures; instead of viewing pictures in a grid, we scatter them on the desktop and allow them to partially occlude one another. When the pictures are ordered in a list, they are positioned in a spiral and stacked in layers so that the first picture in the list is in the center of the screen and in the top layer and the last somewhere at the border in the bottom layer. A picture in a higher layer can partially occlude a picture in a lower layer. All pictures are randomly oriented, as shown on figure 2. To be able to identify which picture the user is looking at, we blur all of them, except one. That picture has focus. As future work, this could be replaced with automatic gaze detection.

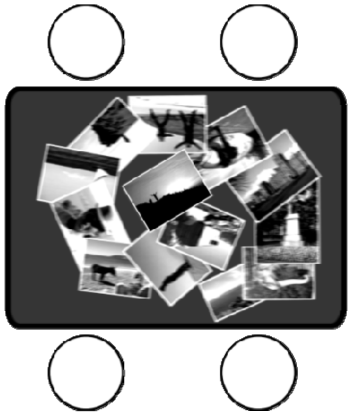


Fig. 2. An example of the visualization of a collection of pictures on a multitouch tabletop display with four users around it. The spiral organization and the random orientation of the pictures puts all the users in an equal position for browsing through the collection.

Although not strictly required, a multitouch tabletop display is used to enhance resemblance to viewing of analogue photos. When more users are sitting around a table, the top left corner of that table for a user is the bottom right corner for another. Placing the pictures in a grid puts users in an unequal situation. This is the reason we place pictures in a spiral with a random orientation.

4 Preliminary Results and Future Work

The above described model needs experiments that will help us evaluate our ideas. First of all we must prove that there is a link between viewing time and interaction with the user's affinity for a certain picture. Additionally, by experimenting with users we will also be able to refine the model and its variables like the weights used in calculating PAS and decay speed.

Our first experiments focused mainly on two things: on validating the main idea behind ShoeBox (viewing time as a measure for the users' affinity for a picture) and on gaining feedback from users and these preliminary experiments themselves on how to design further experiments that will help us to properly quantify the variables of the model. During the experiments we also observed how users interacted with the pictures in order to get some clues on how interaction could be interpreted as a measure for affinity.

Six users were involved in the experiment, three females and three males aged between twenty-five and thirty. After getting accustomed to multitouch interaction they started the experiment in which they were presented 50 photos obtained from Flickr with the keywords “sport car", "flower", "mountain", "winter", "party", "portrait", "war", "landscape", "dog", "see"; 5 pictures for each keyword.

To understand if viewing time is a meaningful measure for affinity for a picture we allowed the users to view the pictures for two minutes. After that, the most and least viewed ten were selected and arranged in pairs; the most viewed picture with the least viewed and so on. The user was then asked which picture he or she likes the most for each pair. We expected pictures with higher viewing times to be chosen.

The users offered a lot of insight on how to plan next experiments, such as to allow them more time for browsing or give them more space to move the pictures around. A user suggested that pictures should implement inertia so that he could swipe away the ones he doesn’t like. Another thing we noticed was that some users spontaneously started stacking pictures in two piles; one with pictures they like and one with pictures they don’t like. This reinforced our assumption that it is possible to extract the user’s affinity for a certain picture not only from viewing time, but also from how he or she interacts with the picture. Despite first getting accustomed with multitouch interaction, some users still experienced problems while browsing pictures.

Besides providing us with guidelines for future work, the experiment also showed that viewing time is indeed related with the users’ affinity for a picture as five out of six users were predominantly choosing the picture with the higher viewing time. Table 1 shows the results for all six users. Overall, in 60 pairs the picture with the higher viewing time was chosen 37 times (61%). Taking into account that the only user that choose more pictures with lower viewing times is the one that experienced the most problems with multitouch interaction, this preliminary results confirm our assumption about viewing time being a measure for the user’s affinity for a picture.

Future work will focus on defining the experiments and conducting them on a larger scale so that the experiments will have real statistical value. Furthermore in the next experiments we will ask the users to bring their own photos to see whether there is a difference in the relation between viewing time and the affinity for personal or general photos. After that we will be able to fully define the affinity model and the focus of our research will shift from the definition of the model to the exploitation of the model in personal information management and in image retrieval tasks.

Table 2. Number of pictures with higher and lower viewing times chosen for each user

User	No. of pictures chosen (higher viewing time)	No. of pictures chosen (lower viewing time)
1	7	3
2	3	7
3	8	2
4	6	4
5	7	3
6	6	4

5 Conclusions

In contrast to currently available software that organizes photo collections according to folder structures, metadata or semantic technologies we provide a method that takes into account the users' emotional aspect by organizing pictures according to his or her affinities, without requesting the user to explicitly express his or her affinities. This is done by measuring the time spent viewing a picture. Also, the organization obtained this way is dynamic, meaning that it changes through time resembling changes in the user's affinities. Furthermore we present an alternative, natural way of visualizing photo collections. As it is desirable of photo managing software to be able to sort photos according to preference, date etc. we see our work not replacing, but complementing current organizing methods.

Acknowledgments. This research was funded in part by the European Union, European Social Fund, Operational Programme for Human Resources, Development for the Period 2007-2013 and by the Slovenian Research Agency.

We would also like to thank our friends who volunteered for experimenting with ShoeBox; for their time spent testing and for valuable feedback on the implementation of ShoeBox itself.

References

1. Picasa, `http://picasa.google.com`
2. Flickr, `http://www.flickr.com`
3. Facebook, `http://www.facebook.com`
4. Latif, K., Mustofa, K., Tjoa, A.: An Approach for a Personal Information Management System for Photos of a Lifetime by Exploiting Semantics. In: Bressan, S., Küng, J., Wagner, R. (eds.) DEXA 2006. LNCS, vol. 4080, pp. 467–477. Springer, Heidelberg (2006)
5. Liu, D., Chen, T.: Unsupervised image categorization and object localization using topic models and correspondences between images. In: International Conference on Computer Vision (ICCV), pp. 1–7. IEEE Computer Society, Los Alamitos (2007)
6. Platt, J.C., Czerwinski, M., Field, B.A.: PhotoTOC: automatic clustering for browsing personal photographs. In: Proceedings of the Fourth IEEE Pacific Rim Conference on Multimedia, pp. 6–10. IEEE, Los Alamitos (2003)
7. Loui, A.C., Savakis, A.E.: Automatic image event segmentation and quality screening for albuming applications. In: ICME 2000, vol. 2, pp. 1125–1128. IEEE Press, Los Alamitos (2000)
8. Graham, A., Molina, H.G., Paepcke, A., Winograd, T.: Time as essence for photo browsing through personal digital libraries. In: Proceedings of the 2nd ACM/IEEE-CS Joint Conference on Digital Libraries, pp. 326–335. ACM, New York (2002)
9. Kang, H., Shneiderman, B.: Visualization methods for personal photo collections: Browsing and searching in the photofinder. In: IEEE International Conference on Multimedia and Expo (III), pp. 1539–1542. IEEE, Los Alamitos (2000)

Toward Adapting Interactions by Considering User Emotions and Capabilities

Idoia Cearreta and Nestor Garay-Vitoria

Informatika Fakultatea, University of the Basque Country/Euskal Herriko Unibertsitatea
Manuel Lardizabal 1, 20018 Donostia, Spain
{idoia.cearreta,nestor.garay}@ehu.es

Abstract. Ambient Intelligence and Affective Computing areas are working on trying to make interfaces more natural from the human side. Taking personal, device and system characteristics into account, adaptive interfaces might help to achieve a greater user satisfaction and it is expected the results can be enhanced when bearing user affectivity in mind. In this paper, an ontology that describes affective interactions has been extended in order to represent information about user's sensorial and perceptual capabilities when he/she interacts with systems. Two use cases applying ontology are presented herein.

Keywords: Affective Computing, Adaptations of Interactions, Ontology.

1 Introduction

Some users who have physical or cognitive disabilities experience difficulties interacting with interfaces; for example, when they want to access the Web or use conventional computer applications. Moreover, most of these applications ignore the specific needs of people and their real capabilities to carry out certain tasks. However, many of these accessibility barriers can be overcome by using systems that automatically adapt user interfaces or the mode of interacting with them. This adaptation can be based on user needs and capabilities and on the context with which the user is interacting.

Furthermore, it is really important for people with disabilities that these applications, especially applications used as communicators, are able to manage the information which is implicitly transmitted, such as speech modulation or facial expressions. In a common human conversation most of the information is transmitted through this kind of stimuli [1]. These stimuli help us to disambiguate the message and properly understand the interlocutor's affective state.

Therefore, it is considered necessary to create models that represent all of this knowledge. Nowadays, one of the most popular mechanisms for modeling the knowledge of a specific domain is the use of ontologies [2, 3, 4]. The main objective of ontologies is to represent concepts of the real world. Thus, it is possible to share the knowledge among people and even among different applications or software agents. Moreover, ontologies provide enough mechanisms to reuse the domain knowledge and extend it by adding new concepts.

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 525–534, 2011.

For structuring the required knowledge and inferring new information from it, an ontology was defined based on a previously created ontology called *Affinto* [4].

This paper is organized as follows. Section 2 reviews related references. In section 3, *Affinto* ontology is presented as it was originally known, and some changes have been incorporated in order to make suitable adaptations in interactions. Section 4 presents two use cases where the extended ontology has been applied. Finally, several conclusions and future works are highlighted in the last section.

2 Related Work

Efforts related to the design and development of adaptive interfaces can be found in the literature [5] applied to human diversity. There are cases where human diversity includes design for all [6, 7]. Recently, the Ubiquitous Computing paradigm made evident the need to take even device characteristics into account for the interface adaptation process [8].

Among the different methods for integrating required knowledge modeling and reasoning mechanisms in adaptive systems in order to make adaptations, the use of ontologies can be found in several references. For instance, Hervás and Bravo present three ontologies in their work (User, Device and Physical Environment Ontology) [8] for describing the world around applications and users. Using these ontologies, they propose an infrastructure to support information adaptability for users. In [7], Abascal et al. present a system for adapting user interfaces to people with disabilities. This system uses an ontology divided into two parts: user capabilities and adaptive interfaces. This work will be explained later since the extended *Affinto* ontology presented in this paper will be applied with it.

In almost all cases the adaptation is done without considering user's affective states. However, emotions affect relationships between people and their surrounding environments, where other people and systems are included [4]. The way to communicate to people and their environment is usually composed of different communication modalities or channels [9].

This paper presents an ontology that, apart from considering the emotions expressed in user-system interactions, takes their communication capabilities into account. Thus, systems provide users with the information transmitted in the most suitable communication modality for them.

3 Ontology for Adapting Interactions

As previously mentioned, the *Affinto* ontology [4] has been extended in order to incorporate information about user sensorial and perceptual capabilities. In an analogous way, the information about the system's capabilities, in respect to modalities of communication, is also incorporated within the ontology. Thanks to the modeling of these new concepts, the most suitable type of interaction for a given user can be inferred.

This extended ontology can be applied to systems that automatically generate interfaces [7]. Thus, the ontology can provide information about both the user and the device's characteristics in order to choose the most suitable multimedia resources. Moreover, it can be applied to multimodal interaction systems. In this case, it can suggest which modality of communication the system should use to interact with a particular user.

Apart from added concepts, the *Affinto* ontology provides knowledge about affective interactions. The reason why new information about communication capabilities was added to an ontology of this type is because it was found essential to include affective interactions for improving naturalness. That is, although the system knows the most appropriate mode of interaction for a given user, if it does not interact in a natural and expressive way with the user, the interaction will remain inadequate for that person.

To this end, Cearreta proposed taking some contextual elements or characteristics into account (subject, environment, socio-cultural, task and spatio-temporal contexts), because they can influence a user's needs and affective states [4]. Moreover, under certain conditions and situations of an environment, it would be appropriate to adapt some characteristics of the interface to generate a more tailored one.

This context model is the basis of the affective interaction definition because it describes the different factors that are necessary to both generate and recognize the affective states of users. In addition, this model allows the coherent integration of above mentioned new concepts, since when using the subject context it is possible to describe cognitive processes.

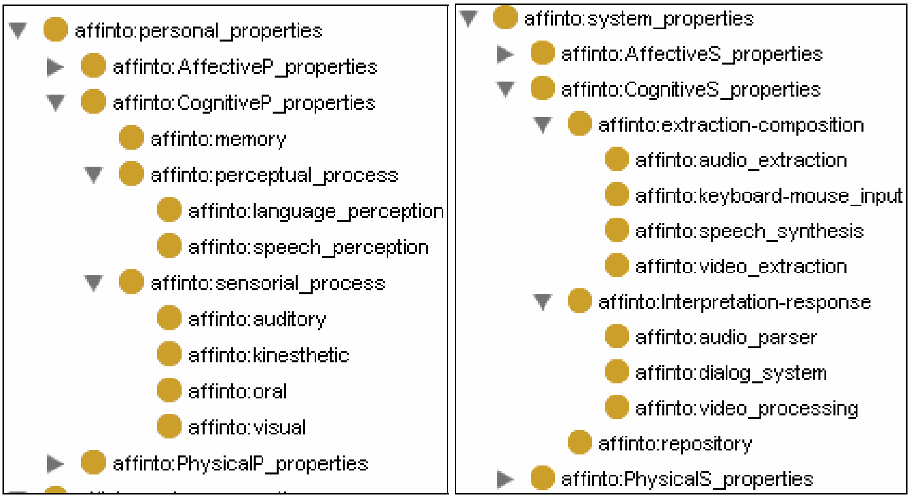
Therefore, new concepts related to the level of communication capability of the user and the system are:

- User sensorial processes (see Fig. 1.a): auditory, oral[1], kinesthetic and visual processes [10], in order to know the user's communication capability levels of these sensorial processes.
- User perceptual processes (see Fig. 1.a): language and speech perception [11], in order to know the user's communication capability levels of these perceptual processes.
- System extraction-composition related concepts (see Fig. 1.b): audio extraction, keyboard-mouse input, speech synthesis and video extraction processes, in order to know the system's communication capability levels of these extraction-composition processes.
- System interpretation-response related concepts (see Fig. 1.b): audio parser, video processing and dialog system, in order to know the system's communication capability levels of these interpretation-response processes.

In addition, new concepts related to the task are (see Fig. 2):

- System service purpose: describes priorities among the communication types of the system. For example, the most suitable modality for emitting or receiving information can be indicated, according to the service provided by the system (see Fig. 3).
- Affective interaction modality: describes modalities that have been used in the user-system interaction (see Fig. 3), after checking compatibilities between user and system capabilities.

[1] Although oral processes are not actually sensorial processes, in this work they will be considered that way.



(a) (b)

Fig. 1. (a) *Affinto's* new concept about user-related sensorial and perceptual processes; (b) *Affinto's* new concept about system-related extraction-composition and interpretation-response processes (analogous to the concepts of user sensorial and perceptual processes)

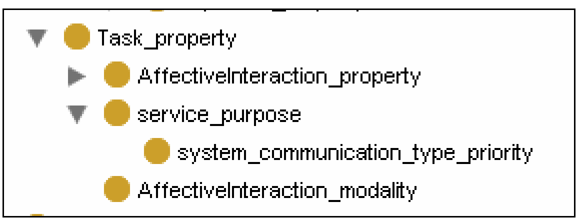


Fig. 2. *Affinto*'s new concepts related to the task

Fig. 3. Communication types used for indicating a system's communication priorities or for indicating the communication types used in user-system interactions

An example is presented to explain the process of applying this ontology. This process could change according to the type of system where the ontology is to be integrated because, as previously mentioned, the ontology can be applied to an automatic interface generation system or a multimodal interaction system. In the first case, the user could use a device in order to access the service provided by the system. In the second case, a system can be directly accessed by the user. This process will be globally explained in order to simplify the example.

First, the user identifies him/herself through a login procedure. In the case of using a device as an access tool, it will also be necessary to identify the user's device. The ontology should have previously stored information about the user. This information could be extracted from the user explicitly, for example, by filling in some forms, or it could be extracted using data mining techniques and by executing an application in order to extract characteristics from the device. Moreover, the ontology could obtain information about the environment by directly asking the user about the scenario where he/she is, or by using some information extraction techniques, for example sensors or GPS mechanisms.

After the identification procedure, the system infers information by using rules and making inquiries to determine the compatibilities between the user's capabilities, the characteristics of the device and the task or service that the user wants to access. Although the communication modalities that a person can use generally do not vary very often, and although the device that he/she uses might be the same, the media resources provided by the system or the modality that the system uses for transmitting the information could be different. Therefore, it is necessary to check the compatibilities in order to adapt and personalize interfaces and interactions in an appropriate way for the user and for the given system. Thus, the system determines which interaction type is the most suitable for a given user, overcoming most of the existing barriers and providing access to anyone who requires its services.

For a suitable adaptation, the media resources that the system offers through its interface and its communication channels should not avoid information about emotions. To this end, by using the *Affinto* ontology, the system can extract information about the features that are required for emotionally enriching the interface, as well as the stimulus that will be emitted to the user. Conversely, the system can also extract information about certain communication modality characteristics of the user in order to recognize the user's affective state by using the information previously gathered in the ontology.

Therefore, creating accessible interfaces by using the most suitable communication modalities is not considered sufficient. For example, if a person needs to interact with an e-learning system through speech, this system should be able to recognize the user's emotions via this communication modality in order to make the communication more fluid and intuitive. Otherwise, the interface would not be fully accessible in that modality.

4 Use Cases Applying the Extended *Affinto* Ontology

Two different systems have been used to validate the extension of the *Affinto* ontology. The first one is a combination of two prototypes created for us in a previous research work [4]. These prototypes were modified for carrying out the adaptation of interactions according to user capabilities. The second one is a system developed by the LIPCNE[2] group of the University of Basque Country [7], and it has been also adapted for use with the *Affinto* ontology.

[2] http://www.kat.ehu.es/s0139-lipcnec/es/contenidos/informacion/gi0163_personal/es_00163_pe/00163_personal.html

4.1 AFFIN Multimodal Conversational System

As mentioned above, the extended *Affinto* ontology has been applied to the prototypes developed in [4]. In this work, two prototypes of a conversational system that use the *Affinto* ontology were created. One of them is unimodal, for transmitting and receiving information from the system through text, and the other one is multimodal, using text and speech. In this paper, an integration of both prototypes and an extension of the *Affinto* ontology was performed in order to adapt the system to a user's communication capabilities.

This conversational system, called AFFIN, in contrast to most of the affective recognition systems that extract the information from an existing large corpus, is based on a personalized recognition process. That is, the information extracted from the context or from the user for applying the affective recognition process to is classified in the ontology based on the interlocutors and their identification numbers. Conversely, when it is necessary to infer information about a certain subject, the instances that belong to that user are extracted from the ontology.

Therefore, being a personalized process, it is necessary to train the conversational system with the user's speech characteristics. For this, the user must carry out several audio recordings, indicating his/her real affective state. However, before starting the training process, the most suitable type of interaction must be determined depending on the user's capabilities. For instance, if the user cannot express him/herself through speech, AFFIN should not provide speech as an input channel. In this case, it will analyze the text that the user will introduce into it.

To this end, after the user's identification, AFFIN asks a set of questions in order to compile the user's profile and infer the type of interaction that the user needs. This inference is obtained by adding rules to the ontology. The Semantic Web Rule Language (SWRL)[3] was used (see Fig. 4) to write these rules in first-order logic.

In order to create a user profile of capabilities, as previously mentioned in section 3, sensorial and perceptual process-related concepts were modeled. At this time, with AFFIN being an approach for adapting interactions according to user capabilities, the questions raised to users are quite explicit and they can be morally intrusive. Hence, indirect questions should be asked, or a set of simple games to determine the user's disabilities or restrictions could be used, without asking them these questions directly and risking hurting their feelings.

According to the user's level of communication capabilities, the interaction that AFFIN provides will be different. Table 1 shows the types of interaction and the combinations of sensorial and perceptual capabilities that AFFIN now supports. If the user has a high level of capability in all modalities, AFFIN gives priority to the speech communication modality since it is a conversational system and the most common way for humans to communicate is through speech. In this particular case, the rule that *Affinto* uses is "interaction_1". As Table 1 shows, "interaction_2" is addressed to users who have hearing problems, especially for post-lingual deaf people. In this interaction type, the alternative which AFFIN offers is to use text as the output communication channel. AFFIN will choose "interaction_3" if the user has oral impairments. For this interaction, AFFIN uses text as the input channel. The

[3] `http://www.w3.org/Submission/SWRL/`

```
<swrl:Imp>
  <swrl:head rdf:parseType="Collection">
    <swrl:DatavaluedPropertyAtom>
      <swrl:propertyPredicate rdf:resource="#reference"/>
      <swrl:argument1 rdf:resource="#USER"/>
      <swrl:argument2 rdf:datatype="&xsd;string">interaction_3</swrl:argument2>
    </swrl:DatavaluedPropertyAtom>
  </swrl:head>
  <swrl:body rdf:parseType="Collection">
    <swrl:DatavaluedPropertyAtom>
      <swrl:propertyPredicate rdf:resource="#oral_capability_level"/>
      <swrl:argument1 rdf:resource="#USER_SPEECH_L"/>
      <swrl:argument2 rdf:datatype="&xsd;string">no</swrl:argument2>
    </swrl:DatavaluedPropertyAtom>
    <swrl:DatavaluedPropertyAtom>
      <swrl:propertyPredicate rdf:resource="#speech_perception_level"/>
      <swrl:argument1 rdf:resource="#USER_SPEECH_PERCEP"/>
      <swrl:argument2 rdf:datatype="&xsd;string">high</swrl:argument2>
    </swrl:DatavaluedPropertyAtom>
    <swrl:DatavaluedPropertyAtom> [4 lines]
    <swrl:DatavaluedPropertyAtom> [4 lines]
    <swrl:DatavaluedPropertyAtom> [4 lines]
    <swrl:DatavaluedPropertyAtom> [4 lines]
    <swrl:IndividualPropertyAtom> [4 lines]
    <swrl:IndividualPropertyAtom> [4 lines]
    <swrl:IndividualPropertyAtom> [4 lines]
    <swrl:IndividualPropertyAtom> [4 lines]
    <swrl:IndividualPropertyAtom> [4 lines]
    <swrl:IndividualPropertyAtom> [4 lines]
  </swrl:body>
</swrl:Imp>
```

Fig. 4. A SWRL rule of the *Affinto* ontology which corresponds to the interaction called "interaction_3", and which is addressed to users with oral impairments (see Table 1 below)

Table 1. Types of interaction and combinations of sensorial and perceptual capabilities supported by AFFIN

Sensorial/Perceptual → Interaction ↓	Oral	Visual	Hearing	Kinesthetic	Language perception	Speech perception
1	high	high	high	high	high	high
2	high	high	**post-lingual**	high	high	**no**
3	**no**	high	high	high	high	high
4	**no**	high	**pre-lingual**	high	**low**	**no**
5	high	high	high	high	**low**	**low**

"interaction_4" rule is addressed to people who have both hearing and oral impairments, especially pre-lingual deaf people. In this case, AFFIN emits and receives information by text. In addition, the text that the system emits will be simplified using a different dialog system [4], since the average reading level of the deaf population is lower than that of the general population [12]. In the above mentioned case "interaction_2" texts are not simplified, since written texts are usually used to communicate with deaf people who can speak, i.e. people with hearing loss or post-lingual deaf people. Finally, "interaction_5" will be selected for users who have difficulties understanding the language, either for a cognitive disability or when the language is not correctly known. In this case, AFFIN will use simpler phrases with a different dialog system as well.

4.2 The EGOKI System for the Automatic Generation of User Interfaces

Extended ontology has been also applied to an automatic interface generation system called EGOKI [7]. According to user communication capabilities, the system includes adequate resources for a given user and applies corresponding transformations in the user interface so that the user can efficiently access the service provided by the system.

Although the EGOKI system has its own ontology with user capability-related and interface-related concepts, for this validation the capability-related part has been replaced. In this way, the aim is to prove its functionality with the extended *Affinto* ontology using similar concepts. The interface-related ontology part has not been replaced, since this includes concepts relating to the resources and adaptations needed for the interface adaptation. The *Affinto* ontology does not describe resources that compose the system interface. What *Affinto* describes is the different communication modalities of both the user and the system as an adaptation mechanism of interactions that are carried out between both interlocutors. Thus, *Affinto* uses SWRL rules, as previously explained (see Fig. 4). These rules are also used for inferring which type of interaction should be carried out by EGOKI, combining different capability levels that a given user could have. In this case, in addition to the five combinations shown in Table 1, new rules have been added (see Table 2) in order to provide the same options that are supported by EGOKI. Once the corresponding interaction with a given user (by means of reasoning with *Affinto*) has been obtained, the EGOKI system makes the query in its own ontology. This query is made in the part where interface resources and adaptations are described. Finally, these resources and adaptations will be applied for creating the interface.

Table 2. Types of interactions and combinations of sensorial and perceptual capabilities that have been added for the EGOKI system

Sensorial/ Perceptual → / Interaction ↓	Oral	Visual	Hearing	Kinesthetic	Language perception	Speech perception
6	high	**low**	high	high	high	high
7	high	**no**	high	high	high	high

As Table 2 shows, the rules added corresponding to “interaction_6” and “interaction_7” are related to the user’s visual sensory processes. According to the user’s capacity level in this modality, the type of interface will be different. If the user has poor vision, the adaptations that EGOKI provides will be to enlarge the text and resources, such as images or button commands, using “interaction_6”. With respect to “interaction_7”, which is addressed to a blind user, EGOKI will generate an interface with plain text so that a screen reader can generate a synthesized voice. In the near future, EGOKI will extract the information gathered in *Affinto* to generate a more expressive voice, thus improving the interactions between the user and EGOKI.

5 Conclusions and Future Works

As previously mentioned, nowadays the use of interactive systems is increasing, and adapting their interfaces and interactions could improve their accessibility. Thus, people with special needs could interact with these systems by using their most adequate communication modalities, also taking emotional transmissions into account. In this work, the *Affinto* ontology has been extended to describe user sensorial and perceptual capabilities in order to adapt and emotionally enrich interfaces and interactions. Therefore, the extended *Affinto* can help adaptive systems in the process of selecting a suitable mode of interaction with users. In addition, ontology can help not to avoid the affective information that the user emits while he/she is interacting with the system by the application of an affective recognition process. Conversely, the system can emit its own emotions, using the information previously gathered in *Affinto*. In this way, apart from exploiting the information about the most suitable interaction mode, a given adaptive system can benefit from Affective Computing technology.

Several examples have been presented in order to apply the extended ontology in real use cases. These examples belong to different types of systems. In the first example, *Affinto* uses some SWRL rules in order to select the most suitable mode of interaction for a given user. Thanks to this information, a conversational system can communicate to the user in a more natural way, tailoring the conversation to his/her needs. In the second example, *Affinto* provides the system with the user’s level of communication capabilities in order to select the most suitable resources and adaptations using the SWRL rules again. In this way, the adaptive system can provide the user with a tailored user interface according to his/her needs.

In future works, an evaluation of these systems will be performed using human experimental subjects as end-users in order to validate the functionality of *Affinto*, after including the previously mentioned sensorial and perceptual capability concepts.

Acknowledgments. This research work was partly funded by the Department of Education, Universities and Research of the Basque Government.

References

1. Mehrabian, A.: Silent Messages. Wadsworth Publishing Co., Belmont (1971)
2. Gómez Pérez, A., Fernández Pérez, M., Corcho, O.: Ontological Engineering: with examples from the areas of Knowledge Management. In: e-Commerce and The Semantic Web. Springer, New York (2003); ISBN: 1-85233-551-3

3. Romero, R.: Especificación OWL de una ontología para teleeducación en WEB semántica. Tesis Doctoral, Departamento de comunicaciones, Universidad Politécnica de Valencia (2007)
4. Cearreta, I.: Affinto: ontología para la definición y creación de sistemas de interacciones afectivas sensibles al contexto. PhD. dissertation, University of Basque Country (2010)
5. Knutov, E., De Bra, P., Pechenizkiy, M.: 12 years later: a comprehensive survey of adaptive hypermedia methods and techniques. The New Review of Hypermedia and Multimedia 15(1), 5–38 (2009)
6. Savidis, A., Stephanidis, C.: Unified design for user interface adaptation. In: Stephanidis, C. (ed.) The Universal Access Handbook, pp. 16-1–16-17. CRC Press, Boca Raton (2009)
7. Abascal, J., Aizpurua, A., Cearreta, I., Gamecho, B., Garay, N., Miñón, R.: A modular approach to user interface adaptation for people with disabilities in ubiquitous environments. Internal Technical Report No. EHU-KAT-IK-01-11 (2011), `http://sipt07.si.ehu.es/icearreta/reports/AModularApproach.pdf`
8. Hervás, R., Bravo, J.: Towards the ubiquitous visualization: Adaptive user-interfaces based on the Semantic Web. Interacting with Computers 23, 40–56 (2011)
9. Lang, P.J.: The emotion probe: Studies of motivation and attention. American Psychologist 50, 371–385 (1995)
10. Cañas, J.J., Waerns, Y.: Ergonomía Cognitiva. Aspectos Psicológicos de la Interacción de las Personas con la Tecnología de la Información, Madrid, Panamericana (2001)
11. Elman, J.L., McClelland, J.L.: Speech perception as a cognitive process: The interactive activation model. In: Lass, N. (ed.) Speech and Language, vol. 10. Academic Press, New York (1984)
12. Barnett, S.: Communication with Deaf and Hard-of-hearing People: A Guide for Medical Education. Academic Medicine 77(7), 694–700 (2002)

A Haptic Emotional Model for Audio System Interface

Yuki Ichiyanagi[1], Eric W. Cooper[2], Victor V. Kryssanov[2], and Hitoshi Ogawa[2]

[1] Ritsumeikan University Graduate School of Science and Engineering
[2] Ritsumeikan University College of Information Science and Engineering,
1-1-1 Nojihigashi, Shiga, 525-8577 Japan
yuki@airlab.ics.ritsumei.ac.jp,
{cooper,ogawa,kvvictor}@is.ritsumei.ac.jp

Abstract. The presented study deals with the problem of selecting music content in digital media, such as mp3 file collections. Usually, to select a specific music file (e.g. a song), one has to directly use some a priori data about the file content, e.g. the artist's name, genre, year of release, or the like. In many situations, however, this data is not visible, does not offer enough information, or otherwise does not provide for any immediately accessible mode for selecting the audio content. With the appropriate models of interaction, haptic output devices have a number of advantages for such selection tasks. First, as haptically enabled systems are becoming common, users are becoming more and more familiar with this modality of user-system interaction. Results of recent studies also suggest that the sense of touch may be more closely associated with moods and emotions than other modalities of interaction. Finally, the sense of touch is available without interference with visual or auditory channels. In the presented study, a model is proposed that links emotional states apparently evoked by music content to specific haptic stimuli. An experiment is conducted to verify tactile-emotive associations assumed by the model, and also to explore whether music specific characteristics, such as genre, would directly be related to haptic sensations. Experimental results obtained are discussed and used to design a novel user interface for an audio system. The envisaged interface would allow for selecting music through tactile interactions. The study's conclusions are drawn, and future work is outlined.

Keywords: User interface, music selection, haptic.

1 Introduction

The demand for online digital music services has increased dramatically over the past few years. For example, in Japan alone, the total sales of digital music content surpassed 767 billion yen (~ $8 billion) in 2009 [1]. Users own and manage large quantities of these digital music files on various devices, and require methods for retrieving the right music files at the right time for full enjoyment. Drawing on a long history of studies about the relationship between cognitive state (typically emotion) and music, researchers have developed systems for music recommendation. Such systems usually rely on user's a priori information and self-reporting of emotional state to infer and

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 535–542, 2011.

retrieve the "appropriate" music. Users, however, are not always able to unequivocally articulate their emotional state, while systems are not always able to readily communicate the emotional content of a musical selection via standard interfaces. Haptic interfaces are a means by which users may be able to quickly and habitually select appropriate music – without having to listen to each selection and without having to pay attention to visual or auditory interfaces [2].

Previous works have established the feasibility of communicating emotional states with haptic cues, including both semantic cues and tactile cues experienced by interaction with haptic force feedback displays [3, 4]. This paper utilizes one of these previous works for application to selection of emotive music via tactile stimuli. A study was conducted to further validate and expand the proposed model by investigating correlations between music, emotions, and tactile stimuli. Section 2 surveys related research. Section 3 presents the tactile emotive association model. Section 4 gives an account of the experimental methods for exploration of music-tactile and also music-emotive associations, and Section 5 provides the results. Section 6 discusses the results and puts forward the concept of a haptically enabled user interface for music selection. Section 7 draws conclusions and outlines directions for future work.

2 Related Work

Affective or emotional state can be hard to convey directly through words or actions. While certain situations and conditions may invoke strong emotive responses, it is sometimes still difficult to express these states. Various cultural, social, and individual differences also confound the ability to readily identify and communicate emotions. One possible mode of communication, inspired by actual human to human communication, is the recognition of emotion in facial expressions. Such studies have often been hampered by the natural human tendency to hide emotional responses under certain circumstances and even to substitute less-genuine expressions [5].

It has, however, been found that haptic and tactile modes of communication are associated with emotional states, and that these associations are common in a variety of societies and personality types [4]. On the other hand, research into the association of music and emotional response has a long history, going back at least to Hevner's definitions of six musical elements and their influence on emotions: key, pitch, rhythm, harmony, and melody [6]. More recently, Balkwill and Thompson showed that even music in unfamiliar tonal systems may be able to reach across cultural and social boundaries to communicate a common emotive content [7]. The authors were also able to identify some specific characteristics of music that influenced emotive content, and proposed a model of psychophysical cues common to all forms of music. Yang et al. proposed a fuzzy classification of the relative strength of emotive content in musical selections [8]. Cai et al. built an "Emotional Allocation" model and a contextual music recommendation based on Web documents which user reads, such as weblogs, and reported a positive correlation between the model and human selections [9]. Trohidis et al. evaluated four algorithms for clustering of emotional content of music as a multi-label classification task [10]. Odagawa et al. [11] developed a music recommendation system for car audio systems in which the user inputs semantic or verbal impressions of the desired emotive content (adjectives such as bright, exciting, quiet, and healing). The system responds with music inferred to match the reported mood. It also incorporates machine learning as well as environmental factors in an effort to improve the recommendation. All these related works may contribute to the development of a music selection interface. So far, however, little has been done to address the problem of intuitive and minimally intrusive music selection.

3 Tactile Emotive Semantics

The model utilized in this paper is based on experimental results of mapping emotional content to tactile stimuli using Russell's circumplex model which were reported elsewhere [3]. Russell's circumplex model classifies emotional state into eight sectors of a two-dimensional space where one dimension is due to dynamics in the cognitive state of alertness, and the other in the cognitive state of pleasure. In the experiments, each sector of the model was identified by a set of semantic descriptors – adjectives suggested by those used in Desmet [12] and translated to Japanese (see Table 1).

Table 1. Emotion circumplex categories and elicitation terms

	Circumplex category	Representative words
1	Neutrally excited (NE)	Surprised, concentrated, eager, astonished, amazed, aroused, longing, avaricious, curious
2	Pleasantly excited (PE)	Loving, jubilant, excited, desiring, inspired, enthusiastic
3	Pleasantly average (PA)	Entertained, admiring, joyful, fascinated, yearning, pleased, proud, surprised, happy, appreciating, amused, cheerful, sociable, attracted
4	Pleasantly calm (PC)	Fulfilled, intimate, satisfied, cozy, comfortable, relaxed
5	Neutrally calm (NC)	Composed, awaiting, deferent, passive
6	Unpleasantly calm (UC)	Gloomy, melancholic, isolated, sad, disillusioned, bored
7	Unpleasantly average (UA)	Contemptuous, disturbed, flabbergasted, jealous, aversive, grouchy, ashamed, cynical, embarrassed, disappointed, dissatisfied, disapproving, confused
8	Unpleasantly excited (UE)	Disgusted, frightened, annoyed, hostile, indignant, alarmed, irritated, frustrated, bewildered, nervous

To select the tactile stimulus perception, the results of Hollins et al. [13] have been used, especially the demonstration that subjective interpretations of tactile surface texture perception can all be satisfactorily described in terms of three orthogonal dimensions. One of these dimensions is, however, composite and breaks down into two sub-dimensions. The adjective scales for tactile perception have thus been defined as "hard-soft", "rough-smooth", "sticky-slick" (or "sticky-slippery"), and "attractive-repulsive". (Note: The Japanese adjectives used in the study may not exactly correspond to the translations given here.).

Fig. 1 displays the model developed through generalization of experimental results reported previously [3]. In the model illustration, double-arrowed arcs indicate the dominant interpretations of the corresponding stimuli and thus specify the basic tactile emotive associations. The model suggests that the surface rigidity-related tactile stimuli tend to be associated with low alertness and low pleasure. In the case of "rough/smooth", the dominant associations are "bored ⇔ smooth", and "unpleasant ⇔ rough". The elasticity tactile perception "sticky" and/or "repulsed" typically provokes unpleasant emotive states. On the other hand, perception of "slick" tends to be linked with relaxation (i. e. boredom), while perceiving attracting forces may induce emotional states related to anxiety. The model also assumes linking softness with pleasant-calm, and stickiness with unpleasant-elevated emotions. As described in the next section, these preliminary results were further used in experiments to examine correlations of music listening, emotive content, and tactile or haptic sensations.

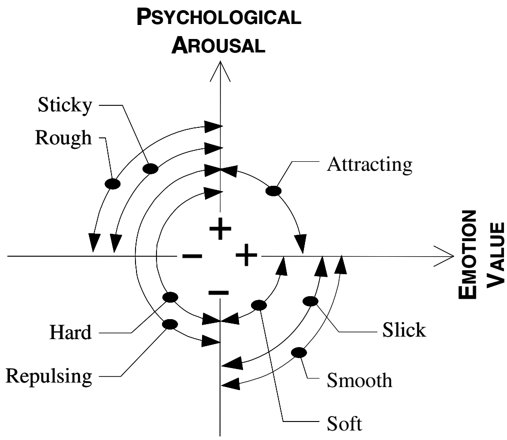


Fig. 1. Semantics of the tactile emotive associations (see [3] for details)

4 Experiment

4.1 Methods

23 subjects (16 male and 7 female) participated in the experiment, all students enrolled in an undergraduate or graduate university program. The average age of the subjects was 22. In order to simplify effects of culture or society in these investigations, the subjects were all Japanese nationals. The music contents used in the experiment were 40 track selections, including Japanese lyrics, English lyrics, and instrumentals. The contents were audited from two GX-70HD ONKYO speakers. The genres of music are categorized as Pop, Rock, Techno, Reggae, and Classical. The selections were intended to be relatively unknown works in order to reduce possible effects from previous listening experience. Subjects were also asked not to speak to other participants in order to reduce cross-subject interference or priming.

4.2 Procedures

Subjects were asked first to listen to the 40 selected tracks of music and then to fill in a questionnaire. Each track was played for at least 1 min. The first question of the questionnaire was used to collect data on tactile associations: "After listening to the music track, please select the tactile perception or perceptions that would best match the music track from the following: A: Hard, B: Soft, C: Rough, D: Smooth, E: Slick, F: Sticky, G: Repulsing, H: Attracting, I: No matching response". The second question collected data on the emotional content in listener responses: "After listening to the music track, please select the numbered emotional category or categories that would best match the music track." The categories given were those shown in Table 1, with the added option of "No matching category". Subjects were allowed to answer with multiple choices. Subjects were also asked to indicate on the answer sheet if they had previously heard the music track.

5 Experimental Results

The results obtained in the experiment are shown in figures 2 and 3. In Fig. 2, the graphs show emotional states selected when subjects listened to every genre of music (i.e. Classical, Rock, Pop, Reggae, and Techno). In Fig. 3, each bar graph describes tactile perceptions selected as the subjects listened to the different genres of music. The black bar represents the statistically dominant response or responses for the given music genre. Tactile perceptions "sticky" and "rough" were selected infrequently in the experiment. There was no significant difference detected between the genders in these results.

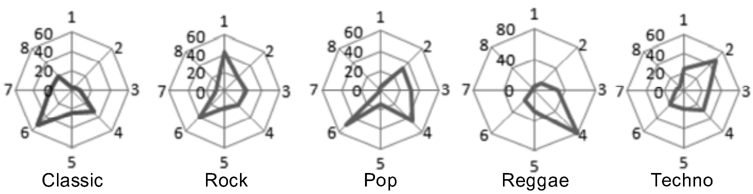


Fig. 2. Music-emotional associations showing total responses for each of the Russell circumplex emotional categories

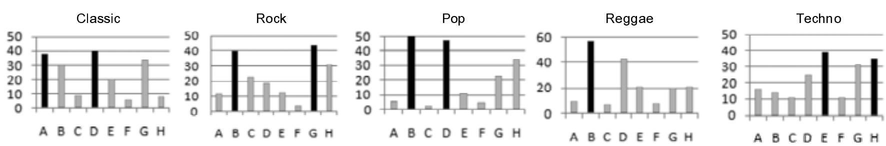


Fig. 3. Music-tactile associations with bars indicating total responses for each tactile perception and black bars indicating statistically dominant responses

6 Discussion

The results of the presented experiment were used to explore tactile and emotional responses to music listening. They first were compared with the model shown in Fig. 1, following the algorithm below:

1. Select music genre and emotional state as presented in Fig. 2. Only statistically dominant results obtained are used for comparison.
2. Use the model given in Fig. 1 to find the tactile perception that best matches the emotional state selected in Step 1.
3. Compare the tactile perception selected in Step 2 with the dominant tactile perceptions obtained as indicated in Fig. 3 for the music genre selected in Step 1.

Results of the comparison are summarized in Table 2. In the table, "○" stands for "the correlation between the model of Fig. 1 and the experimental results is significant at α=0.001," "◎" means "the correlation between the model of Fig. 1 and the experimental results is significant at α=0.05," and "△" indicates that no statistically significant correlation was detected.

Table 2. Comparison of results obtained in the experiment and the tactile-emotive association model

Music genre	Emotion category from Table 1	Tactile perceptions	Evaluation result
Classic	4 (PC)	Smooth, soft	○
Classic	6 (UC)	Hard, repulsing	◎
Rock	1 (NE)	Hard, sticky, repulsing, attracting	◎
Rock	6 (UC)	Hard, repulsing	◎
Pop	6 (UC)	Hard, repulsing	△
Pop	4 (PC)	Smooth, soft	◎
Reggae	4 (PC)	Soft	◎
Techno	2 (PE)	Slick, attracting	◎

As can be seen from the table, the semantic tactile associations established in earlier work [3] are generally supported by the results of the given experiment.

On the other hand, the results revealed no clearly discernible patterns in music-tactile associations. The results in Fig. 3 show that the same tactile stimuli were often linked with quite different music genres. This finding, as well as the established existence of emotive-tactile associations suggests that emotion can be used, rather than genre, as a mediator when one attempts to relay the content of music to the user via tactile stimuli. The apparent music-tactile associations may well be due to user's individual preferences (i.e. specific genre "likeability") which invoke corresponding emotional states (e.g. enjoying pop music while getting irritated with rock). This conjecture leads us to a design of a user interface for selecting music through tactile interactions as depicted in Fig. 4.

In the proposed system, the user can interact with the music selection interface through two modalities. When the user prefers and is able to focus visual attention on the device, she or he interacts with a GUI in a manner common to other media selection interfaces. The system then uses data from this and other interactions to build a model of the user's preferences. User preferences and the tactile-emotive associations described previously determine appropriate force feedback responses as the user navigates the selections available with the haptic interface. Unlike auditory and visual interactions, the haptic sensations are immediate and, after a short period of habituation, allow for intuitive selection of "liked" music with little involvement of the user's attentive cognitive system. The user only needs to initiate and stop the selection scan, based on perceived tactile "representations" of different tracks, where the user's taste is reflected in the tactile stimuli associated with positive emotions.

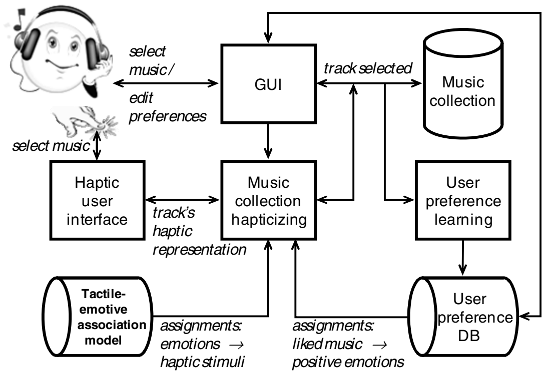


Fig. 4. General outline of proposed interface for haptic music selection

7 Conclusions

This paper described a model to link music and tactile perception via emotion, and put forward the concept of a user interface for music selection in the haptic modality. Data was collected on associations with tactile sensation adjectives and emotional state categories after listening to a track of music. This data and previous results were used to develop and validate a model linking emotional and tactile responses to music listening. The results of this study suggest that human response to music is consistent with semantic associations of emotional state to tactile perception, while the same may not be true for music-tactile associations. In subsequent steps, the authors plan to increase the number of subjects, aiming to obtain more detailed classifications, and further refine and validate the perceptive associations observed. A system prototype is currently under development, which implements the envisaged user interface design, and which would allow the user to select music, based on tactile representations.

References

1. Digital Contents Association of Japan, `http://www.dcaj.org/` (accessed, May 2010)
2. Pauws, S., Bouwhuis, D., Eggen, B.: Programming and enjoying music with your eyes closed. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2000, pp. 376–383 (2000)
3. Kryssanov, V.V., Cooper, E.W., Ogawa, H., Kurose, I.: Computational Model to Relay Emotions with Tactile Stimuli. In: Proceedings of ACII 2009, vol. 1, pp. 288–293 (2009)

4. Hertenstein, M.J., Kelner, D., App, B., Bulleit, B.A., Jaskolka, A.R.: Touch Communicates Distinct Emotion. Emotion 6(3), 528–533 (2006)
5. Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., Taylor, J.G.: Emotion recognition in Human Computer Interaction. Signal Processing 18(1), 32–80 (2001)
6. Hevner, K.: Experimental studies of the elements of expression in music. American Journal of Psychology 48, 246–248 (1936)
7. Balkwill, L., Thompson, W.F.: A Cross-Cultural Investigation of the Perception of Emotion in Music: Psychophysical and Cultural Cues. Music Perception 17(1), 43–46 (1999)
8. Yang, Y., Liu, C., Chen, H.H.: Music Emotion Classification: A Fuzzy Approach. In: Proceedings of the 14th Annual ACM International Conference on Multimedia, pp. 81–84 (2006)
9. Cai, R., Zhang, C., Wang, C., Zhang, L., Ma, W.: MusicSense: Contextual Music Recommendation using Emotional Allocation Modeling. In: Proc. ACM Multimedia, Augsburg, Germany, pp. 553–556 (2007)
10. Trohidis, K., Tsoumakas, G., Kalliris, G., Vlahavas, I.: Multi-label Classification of Music into Emotions. In: Proceedings of the 9th International Conference on Music Information Retrieval, pp. 325–330 (2008)
11. Odagawa, S., Kodama, Y., Gayama, S., Matsushita, F., Suzuki, Y.: A music recommendation system. The Institute of Electronics, Information and Communication Engineers Technical Report 106(160), 49–53 (2006)
12. Desmet, P.M.A.: Designing Emotions, Doctoral Dissertation, Delft University of Technology, Netherlands (2002)
13. Hollins, M., Faldowski, R., Rao, S., Young, F.: Perceptual Dimensions of Tactile Surface Texture: A multidimensional Scaling Analysis. Perception & Psychophysics 54, 697–705 (1993)

Guess Who? An Interactive and Entertaining Game-Like Platform for Investigating Human Emotions

Muneeb Imtiaz Ahmad[1], Hassan Tariq[1], Mehreen Saeed[1],
Suleman Shahid[2], and Emiel Krahmer[2]

[1] Department of Computer Science,
National University of Computer and Emerging Science, Lahore, Pakistan
[2] Tilburg centre for Cognition and Communication,
Department of Information and Communication Sciences
University of Tilburg, Tilburg, The Netherlands
{muneeb06,hassanbintariq}@gmail.com, mehreen.saeed@nu.edu.pk,
{s.shahid,e.j.krahmer}@uvt.nl

Abstract. In this paper, we discuss the design and the development of a highly customizable interactive platform 'Guess Who', which was designed as a tool for investigating human emotions in a variety of experimental setups. In its essence, 'Guess Who?' is actually a game, which includes typical game elements (winning, loosing, scoring) and can also be played purely for entertainment purposes. The design of the game includes three major elements; 1) experimenter-friendly customizable interface, 2) single player mode-laying against an intelligent computer, 3) two player mode - playing with a remote opponent over the internet via audio video communication channel. Early user evaluations show that Guess Who game is not only a very productive research tool for researchers, where researchers are able to record valuable data in different experimental conditions in a natural way, but also a great source of entertainment for children.

Keywords: Games, Emotions, Child-Computer Interaction, Usability.

1 Introduction

Emotions are short-lived multifaceted phenomena [1], which regulate every facet of our daily life [2]. In the past, researchers from many different fields (e.g. philosophy, psychology, sociology, history, etc.) showed an interest in conducting the research on emotions [3]. In the last decade, research on emotions has expanded a lot and entered into many new areas [4] (e.g. neurology, neurophysiology, biology, etc.) and one area where emotions received a lot of response is 'computing'. Here researchers are focusing on making the human-computer interaction more natural and adaptable to human feelings [5].

Despite the fast growth in research on emotions, one permanent challenge faced in the multidisciplinary research on emotion is to elicit emotions in a natural and ethical manner [6]. Traditionally, researchers heavily relied on acted data (where actors were

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 543–551, 2011.

used for eliciting emotions in different modalities e.g. facial expressions, speech etc.) for understanding emotions. It is a well-known fact that acted emotions cannot be a true representation of spontaneous or natural emotions [7]. Furthermore, it has been shown that it is easy to elicit simple emotions using actors e.g. happy, sad, angry but it is generally difficult to elicit more complex emotions (e.g. positive/negative surprise) specially the ones, which are not directly measureable (e.g. shame).

In this paper, we focus on the issue of eliciting emotions in a natural and ethical manner and we primarily focus on four basic emotions (happy, sad, angry and particularly surprise). We present a highly customizable interactive platform 'Guess Who', which was designed as a tool for investigating human emotions in a variety of experimental setups. The design of this platform is built on the idea of the GamE paradigm (Games as a method for eliciting emotions), which has been designed for eliciting emotions in a natural and ethical manner [8].

Games are interactive and dynamic in nature and these attributes make them a good candidate to use as a tool for investigating and eliciting human emotions in different experimental setups specially if the outcome of a game is controllable in a systematic manner [9]. Now a days when games based entertainment industry is rapidly growing, the design of games for serious (educational, training investigation, etc.) purpose is also increasing. Games and emotion research can benefit a lot from each other [10]. On the one hand, games can be used as a tool for eliciting emotions but on the other hand the whole game experience can be enhanced under the 'affective computing' paradigm where games monitor/interpret human emotions and the respond in an appropriate manner for maintaining an optimal level of enjoyment [11].

In its essence, 'Guess Who?' platform is actually a game, which includes typical game elements (winning, loosing, scoring) and can also be played purely for entertainment purposes. The board version of the 'Guess Who?' game is already in the market and is quite popular [12]. Researchers also tried to used it for research purposes but it is very difficult to customize and almost impossible to control the game outcome. We also want to use the computer version of the 'Guess Who?' game as a tool or a platform to collect data from children and investigate two basic questions in a controllable manner 1) how people interact with digital media and show/regulate emotions, 2) how can we elicit spontaneous emotional expressions (mainly facial and vocal) for research purposes. This data can help in understanding human-computer interaction in a better way and will assist in improving our interactive systems e.g. knowledge about different emotional states of humans in different contexts and specially the spontaneous expressions of emotions will lead towards better automatic emotion recognizers (e.g. natural data can be used for training emotion recognizers).

In this paper we not only present the design of the 'Guess Who?' platform in details but we also very briefly discuss the early results from one usability evaluation session.

2 Game Design

In this section we present the design, rules and implementation of the 'Guess Who?' platform.

2.1 Game Rulers

The rules of the game are very simple. Each player has an identical board (blue or red) containing cartoon images of 24 different characters, whose first name is given. Characters may differ along a number of dimensions e.g. gender, color of clothing, amount of facial hair, etc. Before the game starts, each player selects a target card from a separate pile of cards containing the same 24 cartoon characters. The goal of a player is to guess, as quickly as possible, which cartoon character was selected by the opponent. This is done by interactively asking a series of yes-no questions, in an attempt to eliminate candidate characters. Example questions are "Is it a man?", "Does he wear a hat?" etc. After getting an answer from the opponent, a player eliminates various possibilities by flipping down one or more of the 24 playing cards on the board. The player who correctly guesses the opponents character first, wins the game.

2.2 Game Functionality

The software version of the game has three major modules: 1) experimenter friendly customizable interface, 2) single player mode - playing against an intelligent computer, 3) two player mode - playing with a remote opponent over the internet via an audio video communication channel.

Single Player Mode. In the case of single player mode, players play the game against the intelligent computer. When the game starts, a player first chooses the game theme (fantasy, comic, movie, etc.) and after selecting the theme, player chooses a card deck (e.g. Harry Potter, Narnia, etc.). In the figure 1, player chose the 'Classic Guess Who' card deck under the 'Cartoon' theme. After this, the player moves to the next screen where s/he sees the computer avatar (top right corner of the screen) as an opponent. The human player chooses the target card (which his/her opponent has to guess by asking simple questions) from the card deck and that card is displayed on the top left corner of the screen. The system automatically assigns the card to the computer. Once both card are selected, the game counter starts and both players start exchanging basic yes/no questions. The computer uses an algorithm for parsing natural language (NL) and understanding questions. The computer also uses an XML database of card attributes for asking and answering questions in an intelligent manner. The NL interface and the XML database are discussed in the 'Game Architecture' section.

Fig. 1. Guess whose interface with Classic Guess Who deck during single-player game play

Two Player Mode. In the case of two-player mode, two human players play the game against each other disregard of their physical location. Players from anywhere can connect to the server using the given IP address and then choose the opponent. The game follows a client-server modal, which has been used in many online gaming platforms e.g. 'Yahoo Games' where one famous example is 'snooker'. The game uses a client server architecture where server receives all requests from all clients and then dispatches them accordingly. When a user wants to play the game, he/she simply logs into the system using the public IP. A user name and password is required for connecting to the server. Once the user is connected, s/he can choose a game theme and a card deck. For each deck, a player sees a list of available opponents.

In each deck, next to the name of every connected player, three small icons are displayed (video, audio, chat) which show player's preference of communication. If a player does not want to play a game using audiovisual communication then s/he can flag the audio and video icon and other users will understand that the particular player is only available to play a game in the 'chat' mode. Players cannot flag all three options. Players can use different combination e.g. chat + audio, video + chat, etc. During a particular experiment, experimenter can control everything and can force players to play the game according to the experiment requirements. We will elaborate it in the 'customization' section.

Once a player selects an opponent, the main screen is shown. The game starts after both players choose the target card from the deck, which is shown on the top left corner. After each question or decision, players can mark cards for narrow downing their choices. For example if Player A asks a question "is the person male" and Player B replies "No" then player A can check all 'male' cards and these cards will fade out.

The player who correctly guesses the opponents character (card) first, wins the game. It is not possible to cheat in this game. The card, selected as a target card in the beginning, is registered with the system and it is locked for the whole game. At any point during the game play, if any player thinks that s/he is sure about the opponent's card then s/he simply right click on the card and choose the context menu item 'my final guesses. The system compares the 'my final guess' card with the card selected by the opponent and if both cards match, the first player is declared as a winner. An appropriate message is shown on both screens with an invitation to play again.

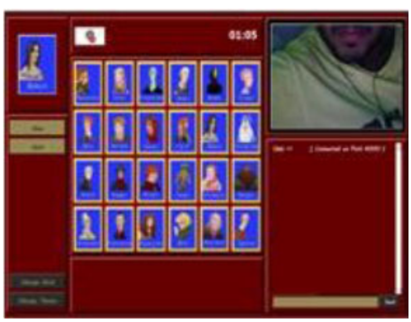


Fig. 2. Guess whose interface with a Harry Potter deck during two-player game play

2.3 Game Customization

The 'Guess Who?' was designed as an interactive and highly customizable platform. In terms of customization, the game had two options: 1) play as an entertaining game and 2) play during an experiment for eliciting emotional expressions. In the first case, it was possible to play this game just as a regular online/offline game purely for entertainment purpose. In this case, no major offline/online customization was required. Users simply play a game either in a single-player or in a two-player mode. We designed this game mainly for children and teenagers and we implemented a number of layout themes. Users had a chance to choose any theme (color scheme of the main interface) at any time.

The second case is much more interesting and for this purpose we designed a customizable architecture. Before the experiment starts, researcher can customize the game based on the research question. Few customizable features are 1) all themes and underlying decks can be changed or new decks can be added immediately. For adding a deck, experimenter needs to place a new folder with a specific name and images at a specific location and the game will immediately load the new deck. A particular order of the cards for each deck can also be ensured by simply following a naming convention, 2) Any stimuli can be used for images e.g. if more complex communication among players is required or researchers want to elicit positive/negative surprise then they can choose characters (on cards) with confusing attributes (cartoonish figure which can be characterized both as a male or female), 3) The game outcome is controllable in the second case. At any point experimenter can change the outcome of the game (e.g. expected winner can lose and vice versa or a player wins a game within a minute because the system showed a clue on the screen (cheated) of one players without informing the opponent, etc.), 4) During the two-player mode, researchers can stop audio or video or both at any time for investigating its effect on communication and expression of verbal/non-verbal cues, 5) The noise of audio channel can be distorted for looking at the effect of hyper articulation in vocal expressions, 6) Fake messages can be inserted in the chat messages for inducing different kind of emotional reactions (only appear on receiver's screen and the sender is unaware of this). 7) Together with the live chat, a prerecorded video can be played, instead of the live video of a player, to see how incongruent or unrelated facial expressions effect the communication. There are number of other ways, in which the platform can be customized. The idea is simple: use one dynamic platform with basic game elements (e.g. winning, loosing, etc.) to understand and answer different research questions. One important thing to note is that all these customization and deterministic features were added based on the requirements and rich feedback of experimenters. The customization rationale is out of the scope of this paper and is not discussed in detail.

2.4 Game Development

We used Python as a programming tool for implementing the game. Python provides a good support for integrating heterogeneous simulation tools in a single environment. Libraries incorporated in this context are "Tinder" for developing graphical user interfaces, "NLTK" for Natural language processing algorithm facilitating single player chat functionality and "PIL" for performing live video streaming (using Video

Capture library as a wrapper). The ‘Guess Who?’ platform is modular in nature. It makes implementation of the text, audio and video chat module easier and reusable. This modular nature also helps in customizing everything on the go. The platform utilizes typical client-server architecture for establishing connection with clients. The application server, which can run together with main server or on a different machine, hosts business logic e.g. game rules and customization principles.

2.5 Game Architecture

The ‘Guess Who?’ platform has following major components 1) XML database, 2) AI agent for single-player mode, 3) centralized communication server, 4) application server (hosts game logic and customization), and 5) application user interface.

XML Database. The XML database is the heart of the platform. The main database contains information about cards and corresponding attributes. In this case, all the decks with related cards and corresponding attributes are stored in an XML file. Every deck has a unique ID and similarly every card in each deck has a unique ID. Each card also has unique attributes depending on the visual features. A card character with a dark complexion who has also long hairs, beard and wearing glasses will have following attributes 1) dark color, 2) beard, 3) long hair, 4) glasses, etc. There is another database of decks, which contains information about all decks available in all themes. Adding a new deck or a new card is very simple. Simply add the card with its corresponding attributes (minimum one attribute is required) and it will appear in the game.

AI Module in Single-player Mode. The single player implementation of the “Guess Who?” game uses a natural language interface for an AI agent capable of playing the game as an opponent. To develop the AI agent, we required a database of intelligent and natural questions, which should be asked by the agent during the conversation. In order to generate this database we recorded the chat sessions of hundred students, who played this game against each other. From the chat records we generated a database of important utterances, which were commonly used by students. The AI agent was developed to ask questions from players about the attributes of different cards. This was achieved by concatenating sentences like “Does he/she have/has”, “Is he/she”, etc., which were the part of questions used in the algorithm, with the attributes (long hair, blue eyes, etc.) stored in the database. In this way a complete grammatically correct English sentence was generated. Similarly the AI agent was supposed to answer questions posted by the human opponent. We used the natural language algorithm, which reads a sentence and breaks it into various parts of speech such as noun, verb and adjectives etc. (NLTK- Brown Corpus). The attributes stored against every card in the database were compared with the attributes of the opponent's question and the system automatically generated a yes/no response. The design of the AI agent is customizable enough to facilitate Multi-language support with the help of NLTK (Natural language toolkit) and XML database of attributes.

3 Usability Evaluation

In the first round of evaluations, overall usability and playing experience of the ‘Guess Who?’ was evaluated. It was important to know if the game can be setup easily for different experimental environments and whether children can play the game without the intervention of experimenters. In this paper we very briefly and in a qualitative manner present the results of our evaluation. The detailed analysis and results will be presented at another venue.

3.1 Usability Evaluation with Experimenters

A usability session was conducted with three participants (experimenters/researchers), two of them were communication scientists and one was a psychologist. All had a good knowledge of computers and experimental design. We designed 3 major tasks for usability evaluation 1) customize the theme and decks (adding/removing cards), 2) setup two computers in two different rooms and run the game after customization and 3) customize the game during live game play. The detailed experimental design and findings of this evaluation will be presented on a different venue. In general, all participants found the platform easy to use and easy to customize. They added one new deck and 3 cards in that deck quite easily. The XML schema was also easily understandable. One user found a major usability problem during setting up the game on two computers. If a webcam is not available in a computer or malfunctioning then the game simply hangs without giving any error message. Actually before initializing, the game checks all hardware components and if it does not find a webcam then it simply hangs. Ideally, it should proceed to the next level after giving a warning. The online customization module specially the interface also needs some attention. Although it serves many general purposes and participants were able to on/off the communication channels during the game play as a task requirement but still it was not easy for every experimenter to use it straight away because it has unique experiment-centric features. Two experimenters showed an interest in having an extra interface for actually customizing the experiment-centric customization features. It is an interesting requirement but it also requires much more flexibility in the core architecture.

3.2 Usability Evaluation with Children

There were two major goals of this evaluation. First goal was to evaluate the game usability and see how much fun children have after playing the game and can they play 'Guess Who?' independently? For measuring this, we used a number of children-friendly methods including fun toolkit [13], mood questionnaires, and pictorial card sorting, and child game experience questionnaire. In this present paper, we report preliminary results of subjective response evaluation.

The second goal was to collect the spontaneous expressions of emotion. For this experiment, the game was customized to induce surprise. It was customized in such a way that 1) children win a game when they were expecting to lose and vice versa, and 2) Change the certainty of winning and losing during the live game play. The reason of this customization was to study the emotional response (facial and vocal) of children to losing and winning situation. For this purpose, game playing sessions were videotaped for further analysis. In this paper, we will not report the experimental design and results of video analysis.

Participants and Procedure. The game was evaluated with 50 children (around 10 years old years) in 2 different primary schools of Pakistan. Both schools use English as a medium of instruction. It was important because children were supposed to communicate with the AI agent using English sentences. 30 children (15 pairs) played the game in pairs (via audio visual communication channel) and 20 children played the game individually (against computer). The game was played in a separate room where a laptop was connected with an external monitor. A camcorder was used for video recordings. In the cases of two-player mode, two separate rooms were used for the experiment. At the start of the experiment, a brief introduction about the game rules was given to the children. The procedure was simple: 1) Children first filled the pre-questionnaire, 2) Children played the game either alone or in pairs and 3) After the game, they filled post-game questionnaires accompanied by a small interview.

Results and Discussion. Early results show that the overall game worked quite well. Children positively interacted with the game in both conditions (playing against the computer and playing against the human player). There were no major usability issues. Most of the kids were able to run the game, select cards, and respond to opponent's questions.

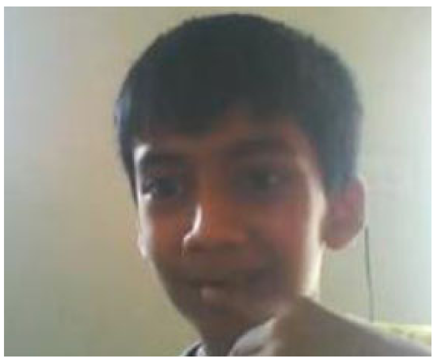


Fig. 3. A child is showing his expressions after losing a game unexpectedly

A number of young children had a problem with using a keyboard for typing but after few tries most of them were comfortable. This was mainly a problem when out of two; one player was slow in typing. In both conditions, the game play had a positive impact on the children's mood and this change in mood was bigger (more positive) in the case of two-player condition. Card sorting results revealed that most of the children wanted to play the game against their friends. There was much richer communication among children when they were playing against friends the duration of the game play was also very long. There were number of problems in the case of single-player mode. One main problem was that the computer was not always able to comprehend the question especially when the right name of an attribute was not used.

We improved it by adding maximum synonyms for each attribute in the database. Disregard of the condition, the top ranked feature of the game was its main rule (guessing) and the audio video communication channel. Overall children found this game engaging, easy to play, full of fun and they were unhappy when the game finished. Majority of children showed a strong interest in playing the game again. Few children did not show an interest in playing again either because they could not understand the game well or they lost the previous game even after a lot of effort. The experiment condition (unexpected winning or losing) had an impact on the game evaluation and those who lost the game unexpectedly (rather wrongly) had less fun and showed relatively less interest in playing again.

4 Conclusion

In this paper, we presented the design, development and early evaluation of a highly customizable interactive platform 'Guess Who?' which was designed as a tool for investigating human emotions in a variety of experimental setups. The results show that the 'Guess Who?' game is not only a very productive research tool for researchers, where researchers are able to get valuable data in different experimental conditions in a natural way, but also a great source of entertainment for children. The early evaluation with researchers also revealed that there is an imperative need of designing new tools for collecting 'spontaneous expressions' of emotion and this platform can fulfill the need in a variety of contexts.

References

1. Lazarus, R.S.: Emotion and adaption. Oxford University Press, New York (1991)
2. Izard, C.: Human Emotions. Plenum Press, New York (1977)
3. Oatley, K.: Emotions: A Brief History. Wiley-Blackwell publishers, London (2004)
4. Gazzaniga, M.S., Ivry, R.B., Mangun, G.R.: Cognitive Neuroscience - the Biology of the Mind, 2nd edn. W W Norton & Company Publishers, New York (1998)
5. Picard, R.: Affective Computing. The MIT Press, Cambridge (1997)
6. Coan, J.A., Allen, J.B.: The Handbook of Emotion Elicitation and Assessment. Series in Affective Science. Oxford University Press, Oxford (2007)
7. Vogt, T., André, E.: Comparing feature sets for acted and spontaneous speech in view of automatic emotion recognition. In: Proceedings of the ICME, Amsterdam (2005)
8. Shahid, S., Krahmer, E., Swerts, M.: Alone or Together: Exploring the Effect of Physical Co-presence on the Emotional Expressions of Game Playing Children Across Cultures. In: Fun and Games, Eindhoven, the Netherlands, pp. 94–105 (2008)
9. Lazzaro, N.: Why We Play Games: Four Keys to More Emotion Without Story. In: Game Developer Conerence (GDC), San Fransisco, USA (2004)
10. Kaiser, S., Wehrle, T.: Situated emotional problem solving in interactive computer games. In: Frijda, N.H. (ed.) Proceedings of the IXth Conference of the International Society for Research on Emotions, pp. 276–280. ISRE Publications, Toronto (1996)
11. Nacke, L., Kalyn, M., Lough, C., Mandryk, R.: Biofeedback Game Design: Using Direct and Indirect Physiological Control to Enhance Game Interaction. In: The Proceedings of the CHI 2011, Vancouver, BC, Canada (2011)
12. Guess Who? Game, http://www.amazon.co.uk/Hasbro-05801-Guess-Who-Game/dp/B00004XQX7 (retrieved on February 2011)
13. Read, J.C.: Validating the Fun Toolkit: an instrument for measuring children's opinions of technology. The Cognition, Technology & Work 10(2), 119–128 (2008)

Adaptive Machine Learning Approach for Emotional Email Classification

K. Karthik and R. Ponnusamy

Department of Computer Science & Engineering
Aarupadai Veedu Institute of Technology – Vinayaka Missions University
Paiyanoor, Chennai – 603 104, India
karthik@avit.ac.in, rponnusamy@acm.org

Abstract. Emotional e-mail classification is one of the important issues in the service oriented organizations. E-mails are served in a first come first serve basis. Few e-mails express the unfair treatment or dissatisfaction of service. It is essential to serve such e-mails with a high priority. In this paper an attempt is made to identify such mails which express the strong emotions of the customers / stakeholders. This system classifies the e-mails in to three categories via positive, negative and other mails. An adaptive machine learning algorithm that uses combined SVD and KNN methods is developed to solve the problem of emotional e-mail classification. Also an emotional dictionary is used as a central component of this system that serves various emotional words and phrases for classification. The system also adaptive in nature and adapts various new words and phrases that explicates the emotion.

Keywords: E-mail Classification, Emotion, Machine Learning Algorithm, Emotional Dictionary.

1 Introduction

Electronic mail is a method of exchanging digital messages from an author to one or more recipients. Modern email operates across the Internet or other computer networks. Today's email systems are based on a store-and-forward model. Email servers accept, forward, deliver and store messages. It is a most preferred communication channel for any service. Educational Institutions are not exception to this. Take for instance, an educational institution which houses around 5000 students and receives the complaints, grievances etc. through email from students as well as parents. Such educational institutions receive on average of 1000 – 1500 emails per month. All emails are served on a first come first serve basis. But, a few e-mails often contain complaints about unfair treatment due to negligence and incompetence. Similarly, a few emails express their high service satisfaction and in such cases it is essential to record messages to measure the value of the service to provide high-quality service.

The emotion fundamentally involves "physiological arousal, expressive behaviors, and conscious experience". Emotion is associated with mood, temperament, personality and disposition, and motivation. These emotions are described by different hierarchy

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 552–558, 2011.

such as primary, secondary and tertiary emotions [1]. Emotions involve complex processes produced by interactions between motives, beliefs, percepts, situations etc. In general the emotions are primarily classified as positive and negative emotions.

In psychology, emotional expression is observable verbal and nonverbal behaviour that communicates emotion. Emotional expression can occur with or without self-awareness. An individual can control such expression, to some extent, and may have deliberate intent in displaying it. A method of expression of emotion is through the verbal communication. Verbal communication is usually understood as the process of communication through sending and receiving word messages. i.e., language is the only source of communication, there are other means also. Written texts have verbal elements to recognize the emoticons.

In this paper an attempt is made to classify the e-mail student and parent messages and emails based on the emotional status. The e-mails are classified into different hierarchical levels as given by Parrott, W. 2001[1]. It contains two levels; at the first level it classifies the messages in to three levels, viz. positive emails, negative emails and other mails. In the case of second levels, it classifies the emails as per the hierarchy.

In this approach we have used adaptive machine learning algorithm that uses the combination of K-NN and SVD algorithms for the classification of emails. Certain words and phrases are already included in the levels of emotional responses, which is helpful in identifying the emotional content of the email. The algorithm introduced is based on adaptive learning, in which the positive/negative emotion n-gram list is learned from the systems previous classification records. An adaptive method can be more efficient, as the emotion n-gram list is modified by the new classification records.

In this paper in Section 2 reviews the previous works in emotional email and message classification. Section 3 explains the components and functions of emotional e-mail classifier in details. Section 4 presents the machine learned classification algorithm in detail. Section 5 gives the implementation as well as experiment results. Section 6 concludes this paper.

2 Previous Works

In 2004, Soo-Min Kim [2] made at attempt to identify sentiments (the affective parts of opinions) from the opinions. This system is able to automatically find the people who hold opinions about that topic and the sentiment of each opinion. Further, Cecilia Ovesdotter Alm, et. al, 2005 [3] explores the text-based emotion prediction problem empirically, using supervised machine learning. The goal is to classify the emotional affinity of sentences in the narrative domain of children's fairy tales, for subsequent usage in appropriate expressive rendering of text-to-speech synthesis. Narendra Gupta[4] and his team, in 2010 made an attempt to present knowledgeable responses to customers' emails are critical in maximizing customer satisfaction. This work reflects customer frustration, dissatisfaction with the business, and threats to either leave, take legal action and/or report to authorities.

Extensive work has been done on emotion detection. In 2003, Liu Hugo [5] and his team had made an attempt to predict basic emotions within text. The authors adopt the notion of basic emotions of Paul Ekman, 1993 [6], and use six emotion categories: ANGER, DISGUST, FEAR, HAPPINESS, SADNESS and SURPRISE.

The work on emotion classification from the point of view of natural speech and human computer dialogues is fairly extensive, e.g. Klaus Scherer's work on vocal communication of emotion [7], Litman and Forbes-Riley's work on text-to-speech synthesis (TTS) [8]. A short study by Sugimoto [9] to addresses sentence-level emotion recognition for Japanese TTS. Their model uses a composition assumption: the emotion of a sentence is a function of the emotional affinity of the words in the sentence. They obtain emotional judgments of 73 adjectives and a set of sentences from 15 human subjects and compute words' emotional strength based on the ratio of times a word or a sentence was judged to fall into a particular emotion bucket, given the number of human subjects. Additionally, they conducted an interactive experiment concerning the acoustic rendering of emotion, using manual tuning of prosodic parameters for Japanese sentences. While the authors actually address the two fundamental problems of emotional text-to-speech synthesis, their approach is impractical and most likely cannot scale up for a real corpus. Again, while lexical items with clear emotional meaning, such as happy or sad, matter, emotion classification probably needs to consider additional inference mechanisms. Moreover, a naive compositional approach to emotion recognition is risky due to simple linguistic facts, such as context-dependent semantics, domination of words with multiple meanings, and emotional negation.

3 Emotional E-Mail Classifier

This system is designed to take the stored e-mails from the server and classify the e-mail in the respective category. The system is designed with the following components, 1. Emotional Dictionary, 2. Classifier and 3. Phrase Extraction Component. These components are explained in detail.

3.1 Emotional Dictionary

It provides a list of emotional words and phrases required to run the system. At the initial stages a list of independent emotional words and phrases are taken from web. The user through the user interface agent to provide systems ontology enters these phrases and words. Later additions of words and phrases will automatically take-place after the identification of new technical phrases. Usually it considers the occurrences of the emotional words and phrase for the new phrase addition.

3.2 Classifier

The classifier initially computes the relationship between the emotional words/phrases as well as the e-mails using the K-NN approach. If it is able to find a match then those e-mails contain the phrases are taken for the analysis. Later, the system arranges the list of words/phrases in the third level sub-category versus list of documents taken for emotional classification. After forming this matrix, the system performs the Singular

Value Decomposing of the given matrix. Then, it computes the correlation between the document and list of words/subcategory. If it finds -1 or +1 then the same document is considered to fall under the emotional category, otherwise it will move to the next category. If it is not able to find the match with any of the given category then the e-mail considered to be the non-emotional e-mail.

3.3 Phrase Extraction Component

This emotional phrase-extraction component performs a list of preprocessing steps. These preprocessing steps involve stop-word elimination, phrase comparison and phrase matching. The phrase comparison through the emotional dictionary is provided with the system. It also does the phrase matching to eliminate slight differences and it will record the new phrases in the emotional dictionary. After extracting the phrase, the system will frame the emotion-matrix.

4 Machine Learning Algorithm

The classifier learns from the set of rules that are passed onto it and based on the level of similarity in the email content and the emotional phrases, further the classifier decides the appropriate Parrot category to which the e-mail should belong. The classifier uses a combined SVD [10] and K-NN Algorithm which is presented below. This choice of machine learning algorithm is not strategic and there is no reason to believe that SVMs or maximum entropy–based classifiers will not perform equally well. The SVD is able to exactly identify the phrase/word relation ship with the words. Also, the number of emotional words and phrases are less and so there would not be any dimensional reduction problem.

The classifier also adds the new emotional words and phrases in the respective categories. In SVD, a rectangular matrix is decomposed into the product of three other matrices. One component matrix describes the original column entries in the same way, and the third is a diagonal matrix containing scaling values such that when the three components are matrix-multiplied, the original matrix is reconstructed. The reconstructed two-dimensional matrix that approximates the original matrix and a few highest values are selected to reconstruct the original matrix.

Each document in the particular sub-hierarchy represents the rows and each phrase with respect to the document is represented as the column. Learning human like knowledge consists in formulating a bi-variate frequency table with row i representing the ith phrase and column j representing the jth document (or between any two entities) and f_{ij} evaluated by the Shannon's measure of information $\sum p \log p$. This together with the dimension reduction will constitute the constraint satisfaction for prediction between the observed and the expected values to make classification. Actual data pertaining to any two measurable entities (phrases and sentences, text classification in digital libraries, etc.) will have to be collected. Sets of examples pertaining to each of the two entities can be exhibited in a bi-variate frequency table for determining the relationships between any two examples.

Machine Learning Classification Algorithm

Input: M - Set of E-mails $\{m_1,m_2,\ldots\ldots m_i\}$, N – Set of emotional words and phrases $\{ n1,n2,n2\ldots n_j \}$, X – Set of emotional E-mails $\{ x_1, x_2,\ldots\ldots x_r \}$

Output: O- $\{o_1, o2\ldots o_r\}$ a set $o_{1=}\{x1_, u\}$ List of emotional mail $x_1$with category u.

Step 1: Begin
Let R = $\{r_1,r_2,r_3\ldots\ldots\ldots.r_i\}$
Let r_1 = List of all words and phrases in message m1, r_2 = List of all words and phrases in message m_2 …r_i list of all words and phrases in message m_i.
Let S = $\{s_1,s_2,s_3\ldots\ldots\ldots.s_M\}$
Let s_1 = List of all emotional words and phrases in message x_1, s_2 = list of all emotional words and phrases in message x_2 …s_r list of all emotional words and phrases in message x_r.

Step 2: Do for k = 1 to i
If ((N x r_i) >=1) then
Add r_i in to X;
End if
End do

Step 3: Do X Not (empty)
For k= 1 to i
A = SVD (N, S)
If (Corr(n1, s_k))>=0.8)
Add s_k to respective N (Change the Emotional Phrase List)
Add (x_k to O)
End If
End For
End Do

Step 4: Procedure End.

5 Simulation Experiment and Evaluation

The system under consideration is simulated using apache e-mail server. In this present experiment the server is hosted with IBM Tehiti server to run the Aglets [11]. The Aglet is a Java based autonomous agent. The Aglets Workbench, developed at IBM's research labs in Japan, is aimed at producing stand-alone agents. The specific emotion agent, phrase-extraction agent and user interfaces agent are developed using these aglets. The standard emotion dictionary developed in this system stores the required emotional words and phrases of each category.

We performed several experiments to compare the emotional e-mail classifier using machine learning algorithm with the combination of K-NN and SVD approach. In these experiments, the Parrot classification system at the three-level hierarchy there exists nearly 135 categories and the present system is trained with few e-mails under each category. To evaluate this approach the server is hosted with sum of 300 e-mails collected form Aarupadai Veedu Institute of Technology e-mails received from parents, students and other suppliers. To evaluate this approach the data sets collected for experimentation are given to designed system and results are compared with Narendra Gupta's work. In order to evaluate the effectiveness of classification system two metrics are measured. The well known metrics Precision, Recall (George Forman 2003) are measured and are shown by

$$\text{Precision} = \frac{\text{Number of E - mails Classified under particular category}}{\text{Total number of E - mails taken for classification}}$$

$$\text{Recall} = \frac{\text{Total Number of E - mails Classified by Machine}}{\text{Total Number of E - mails Classified by Human Expert Judgment}}$$

Table 1. Precision - Recall Comparison

Approach	Precision	Recall
N-Gram Features – Narendra Gupta	0.45	0.61
Combined KNN-SVD Method	0.61	0.72

6 Conclusion

An attempt is made in this paper to design and develop an adaptive emotional e-mail classification using Parrot hierarchy. This work has revealed that the machine learning algorithm using a combined K-NN and Latent Semantic Analysis approach enables one to identify the related emotional e-mails effectively. The initial system is setup with a few set of predefined words and phrases and later on it acquires all the set of phrases. It also yields good results in terms of emotional text identification. This classification system is able to achieve the defined objective of deep level of classification compared to the previous work (Narendra Gupta's) which just identifies the emotion.

References

1 Parrott, W.: Emotions in Social Psychology. Psychology Press, Philadelphia (2001)
2 Kim, S.-M., Hovy, E.: Determining the sentiment of opinions. In: Proceedings of the International Conference on Computational Linguistics, COLING (2004)
3 Alm, O.C., Roth, D.: RichardSproat: Emotions from text: machine learning for text-based emotion prediction. In: HLT 2005: Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing, pp. 579–586. Association for Computational Linguistics, Morristown (2005)
4 Gupta, N., Gilbert, M., Di Fabbrizio, G.: Emotion Detection in Email Customer Care. AT&T Labs - Research, Inc., Florham Park (2010)

5 Liu, H., Lieberman, H., Selker, T.: A model of textual affect sensing using real-world knowledge. In: Proceedings of the 8th International Conference on Intelligent User Interfaces, IUI 2003, pp. 125–132. ACM Press, Miami (2003)
6 Ekman, P.: Facial expression and emotion. American Psychologist 48(4), 384–392 (1993)
7 Scherer, K.: Vocal communication of emotion: a review of research paradigms. Speech Commununication 40(1-2), 227–256 (2003)
8 Litman, D., Forbes-Riley, K.: Predicting Student Emotions in Computer-Human Tutoring Dialogues. In: Proceedings of ACL, pp. 351–358 (2004)
9 Sugimoto, F., et al.: A method to classify emotional expressions of text and synthesize speech, pp. 611–614. IEEE, Los Alamitos (2004)
10 Laham, D.: Latent Semantics Analysis Approaches to Categorization (2003), `http://lsa.colorado.edu/`
11 Lange, D., Oshima, M.: Programming and Deploying Java™ Mobile Agents with Aglets. Addison Wesley Professional, Reading (1998)
12 Forman, G.: An Extensive Empirical Study of Feature Selection Matrices for Text Classification. Journal of Machine Learning Research 3, 1289–1305 (2003)

Designing Poetic Interaction in Space

Yi-Chu Lin[1], Huang-Ming Chang[2], and Rung-Huei Liang[1]

[1] National Taiwan University of Science and Technology, Department of Industrial and Commercial Design, No.43, Sec. 4, Keelung Rd., Da'an Dist., Taipei City 106, Taiwan
[2] Eindhoven University of Technology, Department of Industrial Design, P.O. Box 513, 5600 MB Eindhoven, Netherlands
yihjwu2000@gmail.com

Abstract. An amount of research has risen growing concern of designing aesthetics of interaction in addition to function, usability, and pleasure of it. Beyond aesthetic interaction, we propose poetic interaction as a promising design genre. Based on Bachelard's phenomenological approach, we suggest that poetic interaction design begin with imagination and expression-making of both material and computational things. The challenges include how to design successful reverberation and admiration of poetic images while one interacts and reflects. Examining spatial metaphors, we classify poetic interaction into poetic space and interactive artifacts. With gestalt psychology, we present practical guides for designing poetic interaction. Two examples illustrate the relationship between poetic images and expression-making. Finally, we implement a design work, *whisper*, to explore the framework, as well as validate our findings via a qualitative experiment. Participants were positively impressed that interaction could deliver experience of poetic images with artifacts and space of implicit expression.

Keywords: Interaction Design, Poetic Image, Poetic Interaction, Poetic Space, Expression-making, Gestalt, and Computational Artifacts.

1 Introduction

Over the past few decades, the domain of interaction design has shifted from function, usability, to emotional issues [1] [2]. It seems inevitable that people look for more than functional tools but "living objects." Moreover, some researchers reveal the importance of the aesthetics of interaction [3] [4], the intention of designing the pure beauty of interactions. When the main target is set to fulfill the aesthetics of interaction while designing, the essentials would not be how powerful the function is, or how easy it is to be used, but how the pure feelings or meanings are delivered during the interaction process. Recently, an amount of research reveals the intrinsic value of interaction, which focuses on how users perceive and comprehend the meaning from the interaction with interactive artifacts or within space. Moreover, some studies indicate the implicit qualities of interaction, such as gestalt [4], fluency [5], etc., which help clarify the essence of interaction. Therefore, we wonder what category of interaction design can be further explored beyond aesthetic issues, and what methodology can be applied as guides for such a category.

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 559–568, 2011.

1.1 Poetic Interaction

In terms of Jordan's hierarchy of consumer needs (Fig. 1), an increasing amount of research is toward designing aesthetic interaction that is beyond the hierarchy. Referring to Heidegger's assertion that poetry is in some way the model of all other art forms [6], we propose poetic interaction as the origin and abstraction of aesthetic interaction, which are higher needs according to Jordan's hierarchy. One significant change is that aesthetic appealing is not necessarily valid for poetic interaction. Moreover, poetic interaction design focuses on conveying an image indirectly with an objective expression that contains some uncertain part of it, which causes flashbacks originated from personal experiences. Of notice is that the uncertain part of the expression is the key to convey poetic image in interaction. The delivery of poetic image depends on participants' inherent ability of gestalt. Inspired by aesthetics of interaction, we see a promising aspect of design, which is to create an implicit or transformed expression that invites users to interpret with common sense or their own experiences, and thus, makes users participate and reflect. Therefore, in this paper, we review theories of Bachelard's phenomenological approach, cognitive psychology and design cases, discuss potential ways of conveying poetic image, and establish principles of poetic interaction design. Finally, we implement a design work to validate our findings.

Fig. 1. A hierarchy of consumer needs

1.2 Reflection and Emotional Awareness

Hallnäs and Redström [7] propose Slow Technology, a design agenda for technology aiming at reflection and moments of mental rest rather than efficiency in performance, which leads a way of thinking that works could arouse the awareness and reflection after the interaction process. However, slow technology focuses on sense of time. This paper proposes a new design agenda that could evoke not only users' reflection of personal experience but also an emotional tendency. Aesthetic interaction design cares about the aesthetic experiences while interacting with an interactive system, including not only how it looks but also how it feels [3]. We argue that poetic interaction design inherits the attempt from aesthetic interaction design genre. Kolko [8] has given an initial definition of poetic interaction as "one that resonates immediately but yet continues to inform later, causing reflection and relying heavily on a state of emotional awareness." According to this, poetic interaction would emphasize the following reflection and the emotional awareness rather than the immediate feelings about aesthetics.

According to Bachelard [9], five senses and the existing experience are the bases of "material imagination", which help people to form the latent meaning behind the appearance. Moreover, the image that is formed by material imagination would make people more sensitive of what they feel for the current circumstance. To design for space that arouse reflection and emotional awareness, the importance of expression of computational things needs to be promoted [10]. Therefore, poetic interaction design should begin with expression making and imagination of both matter and computational things.

1.3 Gestalt Psychology

Gestalt Psychologists indicate that human beings tend to see things as a whole rather than see them separately [11]. Moreover, if the parts that comprise the whole are independent, the whole is not simply the sum of the parts, but a synergistic "whole effect", or gestalt [12]. For example, in Figure 2A, people would tend to recognize the image as a white triangle upon three complete black circles instead of three independent incomplete circles and a right triangle (Fig. 2B) or three independent incomplete circles (Fig. 2C). What needs to be noticed is that spatial relationship would be taken into consideration spontaneously through this plane graphic, which indicates that gestalt does not equal to the total of the parts.

Fig. 2. An example of Gestalt psychology

As Kenya Hara [13] has discussed the concept of "Emptiness", he argues that communication happens when the recipient, offered not a message but an empty vessel, supplies the meaning himself. According to Kenya Hara, the perspective of emptiness is closely related to the poetic image of Bachelard, who points out that the way of conveying poetic image should orient oneirism (dream-like experiences) rather than accomplish it. As a result, under the frame of Gestalt psychology, the notion of the emptiness and unaccomplished part offered a sounder theoretical basis for poetic image. The experience of emptiness is common human experience. Its nature is perhaps best conveyed through poetic image, novels, or the visual arts [14]. With the inherent ability of gestalt and emptiness, poetic image in interaction can be designed by the expressions that are intentionally made incomplete.

1.4 Retentissement and Admiration

The nucleus of the delivery of poetic image mentioned above depends mainly on participants' innate ability of gestalt. However, regarding the interaction design, we think that the perception of poetic image is highly pertinent to harmony in interaction that is inseparable from admiration. In the view of Bachelard [9], "the poetic image is an emergence from language, and it is always a little above the language of signification. By living the poems we read, we have then the salutary experience of emerging." It is poetic image that puts language in a state of emergence. In other words, we have to experience its moment and feel admirable instead of any critical considerations. Hence, extended from the viewpoint of literature, the perception of poetic interaction can be established. There are two points worth noting. First, concerning poetic interaction design, how could poetic image emerge while practicing interaction? On the other hand, how could we create and experience a moment of retentissement (reverberation) and admiration of poetic images after interaction?

2 Design Approach

Lakoff and Johnson differentiate three types of metaphors: structural, orientational, and ontological [15], where structural are metaphors of conceiving of one concept in terms of another and the other two are space related. Orientational metaphors convey one concept in terms of physical orientation based on the fact that we are embodied in physical environment. On the other hand, ontological metaphors give abstract things a sense of substance, allowing us to speak of them as objects. The reason why orientational and ontological metaphors form spatial metaphors of understanding abstract notion is because we have universal experience of being in physical environment and manipulating physical objects. To convey incorporeal and abstract things such as poetic images, interaction design shows promising potential by realizing these two spatial metaphors. Therefore, we classify poetic interaction into two categories: poetic space and poetic interactive objects. No matter which category, we think that expression-making for poetic interaction is the key step toward successful interaction design conveying poetic images. These two categories are illustrated in detail after examining expression of interaction design.

2.1 Expression-Making for Poetic Interaction

Extended from Norman's Conceptual Models [16], the essentials of poetic interaction design can be deployed (Fig. 3). We argue that poetic image and reflection in a participant's mind are the results of reciprocity between mental model of a participant and designers' expression. The context where the interaction takes place provides the "interaction order" among the artifacts and a participant [17]. Of special notice is that the center of expression that evokes poetic meanings is the removed part as well as the visible context.

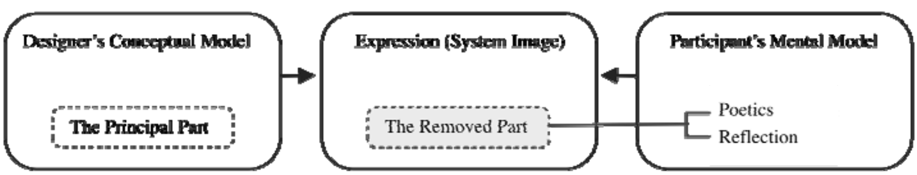


Fig. 3. Conceptual Models of Poetic Interaction Design

To accomplish the implicit part, two questions need to be answered: what should be removed, and how much proportion of the whole expression should be. The answers suggest the gestalt that participants would generate from the interaction. To create expressions for poetic interaction, the procedure is as follows (Fig. 4).

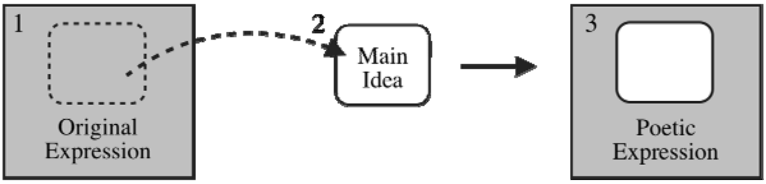


Fig. 4. The expression of a poetic interaction

Step 1. The main element of the designer's concept should be removed from the expression as blanks for participants to fill in. Designers' initial expression might be straightforward at the beginning. Yet to transform the original one to an expression for poetic interaction, the most compelling part needs to be recognized and removed.

Step 2. If some of the peripheral parts were deleted instead of the principal part, the expression would be regarded as an unfinished expression rather than a poetic expression. The proportion of the removed part of the expression would affect the level of implicitness. The expression represents the information that the design work provides, which means that lower proportion of the removed part of the expression would lead to a more explicit expression. Although participants can effortlessly interpret it, the interaction will not be poetic since no space is provided for reflection.

Step 3. If more than the bottom line of the expression is removed, the expression would be inadequate to be reassembled as a gestalt, where participants would not get enough the context and hardly retrieve the poetic image (Fig. 4). Therefore, designers should explore a preferable proportion between implicitness and explicitness.

From the above discussion, at the expression-making stage of poetic interaction design, three essentials are established. First, the context provides most of the background knowledge to the interaction. Second, the principal part of the expression needs to be recognized and removed as a blank for participants to fill in. Finally, the level of implicitness results in the level of poetic image. It is the responsibility of a poetic interaction designer to discover a preferable boundary of implicitness and explicitness. Throughout this, an expression of poetic interaction is considered accomplished.

3 Poetic Interactive Space and Artifact

Interactive space refers to space embedded with computing technologies where interaction is occurring. In such space, if the aim is to deliver poetic image rather than function, usability, pleasure, or aesthetics, we call it poetic interactive space. In addition, in order to design poetic interactive artifacts, it is necessary to find the concepts that belong to the same collocation can be regarded as making metaphors. As argued by Bachelard, "the poetic image is an emergence from language," we think that alternating collocation would be a good way to put language in a state of emergence. Not only changing ways of verbal description of movement, interaction design shows great potential to make users live the poems by acting with emerging collocations, which is carefully selected to form a poetic image.

3.1 Related Work: Office Live 4πversion

Office Live 4πversion [18] is an interactive artwork that attempts to describe a phenomenon about a group of colleagues working together in their own seats. However, this artwork only consists of a few chairs, a table, and some glasses on it. The only interaction that is happening is part of the screen of a colleague's computer would be projected to the glass that represents his attendance to the meeting as if he is making a speech to others. The spatial relationship has implied that groups of people are gathering. When a participant takes a seat and moves a glass, one is interacting within the space while becoming a part of expression. Although the expression is incomplete with the missing, audiences are able to recognize their interactions with their ability of gestalt. In this example, a poetic interactive space invites participants to interactively become a part of expression that is intentionally removed by designers. By acting as the missing part, the main idea, a participant would experience a moment of reverberation and a sincere impulse toward admiration of poetic images, as asserted by Bachelard as the key elements of poetic image.

3.2 Related Work: Tech Tap

Tech Tap [19] is a light. The form of the mainframe is a faucet, which also represents the switch of it. The form of the light is a bottle that can be loaded with light. If a user wants to bring the light with him to another room, he could turn on the light as tapping water, and load the bottle up with light. The light in the bottle would evaporate with time until it is drained. Some of the collocations of light and water are common, such as flow, and hence it is reasonable to apply the image of water to light. Therefore, the original expression is a simple and direct metaphor: light is water. However, as a poetic expression, the water is removed from the expression. Although no real water flows out, the light starts to shine, which correlates to the form of the artifacts smoothly. The ability of creating gestalt enables users to make the connection spontaneously. By literally describing the interaction, poetic expressions would emerge such as "pouring oneself another bottle of light," and "the bottle of light evaporating in the living room."

4 From Theory to Practice: Whisper

By reviewing several design cases, three general principles of poetic interaction design are preliminarily made: (1) the context provides major knowledge for poetic interaction; (2) the principal part of the expression should be removed; (3) to perform in a level of implicitness, intangible materials are preferably used to deliver poetic image, such as sounds and light. To exemplify the principles, a design work is done: whisper. From observation and interview, a drawer could be seen as a personal space to store and conceal precious things. Moreover, the behavior of finding things in a drawer is similar to that of digging things in the soft ground. Inspired by the well-known fairy tale "the King's Donkey Ears", expressions are made: everyone has a drawer that is connected to the Internet, like a square of shared ground; users could bury their secrets in their own drawers; secrets would be synchronized automatically; the finder is the secret keeper; buriers and finders are anonymous. The shape and function of the drawer remain the same, keeping the artifact in daily use and preserve the original context of the interaction between drawers and users.

4.1 Concept

According to the above expressions, the digging and burying movements are of special interests (Fig.5). When one hides things in the drawer, it can be interpreted as burying treasures in the ground. With this metaphor, this work invites users to find an empty space in it, tell their secrets as if they are in a wilderness, and record it as a *whisper* (Fig.6). After finishing recording, the embedded LED light at the bottom of the drawer will be switched on to indicate its spot, and the objects need to be put back to the initial position to bury the secrets.

Fig. 5. Digging things in the ground

Fig. 6. Burying things as a whisper

When a whisper is buried, all drawers connecting to this system would be synchronized, implying that all drawers share the same mysterious land containing whispers. Therefore, when the drawer is opened, it will play the sound of nature to create calm and relax atmosphere. When another user finds a whisper, it will play the recorded voice only once, which means "takers are keepers". Meanwhile, the LED will be off as if the whisper is taken away. All the communications proceed through Internet. While a whisper is buried or taken away, a tweet will be posted on twitter to make a history of this ongoing interaction for users to trace. This concept conducts an exploration of applying everyday objects to realize the vision of the Internet of things.

4.2 Poetic Interactive Spaces and Poetic Interactive Artifact

In this work, poetic interactive space and artifact are harmoniously integrated. A drawer, as a private and intimate space, contains personal tiny objects, which might be put casually or hidden carefully. The sound of nature played while opening a drawer refers to the space as a wilderness at night. With the context of fairy tales implied by the sound and the impression of its spatial meanings, the expression of the space is established. As we have mentioned about poetic expression, there are no trees or soil put in the drawer as the main part of the concept. On the other hand, the bottom of the drawer, the *whisper* recorder and player as well as personal items can be regarded as interactive artifacts. The metaphor is a land with holes that contain *whispers*. However, physical holes and substances are removed from the expression. Therefore, the expression of poetic interactive artifacts is accomplished. Moreover, literally describing the interactive experience with *whisper* yield poetic expression such as "burying one's voice in a drawer," or "digging one's secret serendipitously."

4.3 Experiment

A qualitative experiment is held to investigate the interactive experience of the participants. The main purpose of this experiment is to determine if there exist any subjective feelings evoked, by collecting adjectives that participants use to describe this work. The participants were 20 students from a wide range of academic areas, ranging in age from 20 to 35 years. For each user, the procedure includes 3 steps: interviewing for user experience, playing the scenario, and exchanging and sharing.

First, to understand the background of participants, they were interviewed about their life experiences and habits, etc. Furthermore, chatting in this space would elicit information concerning one's attitude and situation. The purpose of this step is to immerse in the context of experience rather than finish a task of experiment.

Second is playing the scenario**.** At the beginning, we inform each participant about the behaviors of digging and burying something briefly instead of offering too much detail information. Each can hear secrets of others anonymously via drawers. In addition, we invite participants to bury a secret as a sharing procedure. We video-recorded all the experiment processes as well as personal emotions and feedback.

The last step is exchanging and sharing. We gathered participants to discuss and share feeling, thinking, and suggestions. Overall, several situations are of interest. For example, we received an email from one of the participants at that night, saying that he really wanted a *whisper*, which implies that reverberation and reflection after experiment did happen. The poetic images experienced by this participant did emerge as a mental need to revisit the poetic space by designers. On the other hand, although the main element of the designer's concept is removed from the expression, it is surprising that one participant guessed the concept of *whisper* as the tale " the King's Donkey Ears" which was the intended expression we had removed.

4.4 Results

Without explanation in advance, most users felt surprised when opening the drawer that played the sound of nature to create calm and poetic atmosphere. Some of them felt curious and expected about where the voice messages came. Some of them intended to enjoy the variety of the ambient light, etc. By listening to these shared secrets, most participants recalled their personal memories, even including singing a love song. Each participant aroused a private space at a different moment. Moreover, based on the description and sharing of participants, we analyze the empirical data and illustrate our findings using quotes. Three research findings are summarized below, ambience, behavior, and communication.

Ambience. This finding suggests that dynamic interaction between elements in space be inextricably linked with the aesthetic experience. Participants who get involved in this space might feel the poetic imagery by immersing the ambience. This can be seen in the following quotes.

Subject #04: "There is an ambience of nature when I open the drawer. As far as this space is concerned, it can set my mind at ease. Moreover, I immerse myself in this space-time." Subject #09: "I am interested in the shimmer which is buried in the bottom. It is a good idea to make implication of secrets through the shimmer." Subject #11: "I feel imaginary about the interaction. Especially, the sound of nature that plays while opening the drawer, it is so poetic." Subject #14: "I like the variety of the LED when the LED is switched off gradually as if the whisper is taken away."

Behavior. This finding implies that the aesthetic experience of an object cannot be understood without its context and intrinsic values. By interacting with the drawer, participants can experience the poetic image with a reflection of the behavior.

Subject #01: "I like the behavior of finding something in the drawer. To me, it is an intuitive behavior." Subject #06: "It is so interesting to select a shiny secret and makes me wonder what the secret is." Subject #08: "I think it is great to open the drawer while finding a secret accidentally." Subject #09: "The behavior is just like that secrets can be buried in the drawer." Subject #19: "I may open the drawer at anytime because I want to know whether a new secret comes in."

Communication. The essential of the *whisper* is an everyday object. To endow such an artifact with meanings, the purpose of our interaction design agenda emerges. Nowadays, a drawer plays a role that can communicate with each other through the technology, echoing the coming age of IOT (Internet of Things). The anonymous communication adds a poetic interactivity unexpectedly.

Subject #07: "I strongly prefer to communicate secrets with anonyms. I can create an illusion such as appearances or anything else." Subject #13: "It is fascinating that we can guess the teller of the secret implicitly. In addition, we can develop friendship and discuss more topics from the secrets." Subject #18: "I may expect who is listening to my secret, or whether my secret may convey to the people I expect." Subject #20: "It is interesting to guess where the secrets come from. However, *whisper* maybe can create a chatting topic such as do you play the drawer?"

5 Conclusions and Future Work

The present study is initial research on an approach of interaction design delivering of poetic imagery. We believe that there exist various kinds of approaches to implement poetic interaction. Furthermore, it would be a promising direction to establish a framework of poetic interaction. In addition, according to the experiment responses and observation, we feel that the quality of the implicitness has a significant effect on the level of the poetic imaginary. However, the discussion of implicitness is beyond the scope of this paper. Nevertheless, it is worth of noting that how implicit the interaction gestalt is corresponds to the level of the poetic imaginary, creating a spectrum of interpretation. However, too implicit meaning of interaction gestalt makes it difficult for participants to perceive or understand. Moreover, this finding points to an aspect in the spectrum of implicitness and explicitness for future research.

References

1 Norman, D.: Emotion & design: attractive things work better. Interactions 9, 36–42 (2002)
2 Jordan, P.W.: Designing Pleasurable Products. CRC Press, Boca Raton (2002)
3 Hashim, W., Noor, N., Adnan, W.: The design of aesthetic interaction: towards a graceful interaction framework. In: ICIS 2009 Proceedings of the 2nd International Conference on Interaction Sciences: Information Technology, Culture and Human, pp. 69–75 (2009)

4 Lim, Y.-k., Stolterman, E., Jung, H., Donaldson, J.: Interaction gestalt and the design of aesthetic interactions. In: Proceedings of the 2007 Conference on Designing Pleasurable Products and Interfaces, DPPI 2007, pp. 239–254 (2007)
5 Löwgren, J.: Fluency as an experiential quality in augmented spaces. International Journal of Design 1, 1–10 (2007)
6 Heidegger, M.: Basic Writings: from Being and Time to The Task of Thinking. Harper-Collins, New York (1993)
7 Hallnäs, L., Redström, J.: Slow technology–designing for reflection. Personal and Ubiquitous Computing 5, 201–212 (2001)
8 Kolko, J.: Thoughts on Interaction Design: A Collection of Reflections (2011)
9 Bachelard, G.: The Poetics of Space Beacon Press (1994)
10 Hallnäs, L., Redström, J.: From use to presence: on the expressions and aesthetics of everyday computational things. ACM Transactions on Computer-Human Interaction (TOCHI) 9, 106–124 (2002)
11 Lehar, S.: Gestalt isomorphism and the primacy of subjective conscious experience: A gestalt bubble model. Behavioral and Brain Sciences 26, 375–408 (2003)
12 Hara, K.: Designing Design. Lars Muller Verlag (2007)
13 Hothersall, D.: History of Psychology. McGraw-Hill, New York (2003)
14 Hazell, C.: The Experience of Emptiness. AuthorHouse (2003)
15 Lakoff, G., Johnson, M.: Metaphors We Live By. University Of Chicago Press, Chicago (2003)
16 Norman, D.: Emotional design: why we love (or hate) everyday things, 1st edn. Basic Books, New York (2005)
17 Scollon, R., Scollon, S.W.: Discourses in place: language in the material world. Routledge, New York (2003)
18 Office Live 4πversion. Techart Group (2006), http://www.techartgroup.com/
19 Tech Tap. PEGA D&E (2008), http://www.pegadesign.com/

Spectral Subtraction Based Emotion Recognition Using EEG

Jin-Hong Min, Hyeong-Oh Kwon, and Kwang-Seok Hong

School of Information and Communication Engineering,
Sungkyunkwan University, Korea
kaysi@skku.edu, oya200@skku.edu, kshong@skku.ac.kr

Abstract. In this paper, an EEG-based emotion database was reconstructed using spectral subtraction, and recognition performances were evaluated. For subtraction, we created two types of databases. One database included facial expression readings, and the other included both emotion and facial expression readings. A reconstructed database containing pure emotional information was achieved by spectral subtraction, and compared with the original recorded data of emotion and facial expression readings. Facial expression illustrations and the International Affective Picture System were used for inducing facial expressions and feelings. EEG data was recorded after an emotion was excited or while imitating a particular facial expression. By subtracting the database of information related to facial expressions from a database about facial expressions and emotions, pure information about emotion was created. The method used to separate emotion and expression in a database was spectral subtraction. Recognition experiments were classified into six types of emotions. Using the original database, the true emotion could be guessed from EEG readings 29.9% of the time, but using the reconstructed database resulted in an 81.7% recognition rate.

Keywords: EEG, Spectrum Subtraction, Emotion Recognition.

1 Introduction

Interest in user interfaces (UI) is increasing. In accordance with these requirements, human computer interfaces (HCI) have attracted attention as a promising technology on which to base UI. Through HCI technology, users can be provided with a comfortable life in which they have more free time. Human voices, images, gestures and other bio-signals are used for HCI. Today, brain computer interface (BCI) development is also an active field of research. In the BCI research field, brain activity is measured by various methods such as electroencephalograms (EEG) or functional Magnetic Resonance Imaging (fMRI) for emotion or intent recognition. Among these studies, EEG-based emotion recognition has become a relatively new field in HCI. Despite the difficulty in defining emotions, studies about emotions as a part of natural human communication have been conducted. When suitable services for user experiences are provided based on emotion recognition results, user satisfaction is increased.

There are several studies concerned with emotion recognition from EEG signals. In Schaaff et al., a subject-dependent system for emotion classification based on

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 569–576, 2011.

frequency band characteristics is explained [1]. The study of EEG-based emotion recognition is done using the statistical feature of EEG [2]. In Kwang EunKo et al., a feature vector based on the power spectrum of EEG signals is used [3]. However, the main problems of the EEG-based emotion recognition system are the methods of data collection and blind source separation. Thus, the performance of recognition depends on the way the database was collected. If there is a lot of useless data, or noise, in a database, this leads to a decrease in performance. In Junfeng Gao et al., a blind source separation algorithm for the effective removal of eye muscle artifacts from EEGs is outlined [4]. Also, many studies talk about how to recognize emotion without using the Delta frequency band because the movements of muscles influence the Delta frequency band [5]. Some researchers only pay attention to Alpha and Beta bands, which are caused by emotional activities, eliminating the Delta band with a band pass filter. However, this approach leads to a loss of information which was in the filtered band.

This study presents an EEG-based emotion recognition system using spectrum subtraction. When people feel emotions, they also change their facial expressions [6]. However, facial expressions are unnecessary information for EEG-based emotion recognition systems. The artifacts of facial expression can be removed if their characteristics are known. So, we subtract the artifacts of facial expression through the means of a spectrum. We create a pure emotion database by subtracting the mean of the facial expression spectrum from the emotion and facial expression database. Recognition rates are evaluated using the created database of pure emotion.

The structure of this paper is as follows. Section 2 presents a method of data acquisition and spectral subtraction. The recognition experiment and results are described in Section 3. Finally, Section 4 concludes the paper.

2 Data Acquisition

The QEEG-8 records a maximum of 8 channels. So we recorded EEG data from 8 positions based on the International 10-20 System of Electrode Placement (Fp1, Fp2, F3, F4, T3, T4, O1 and O2). The International 10-20 System of Electrode Placement is illustrated in Fig. 1 and the positions that were used in the proposed system are circled in red. The sampling frequency was 256Hz.

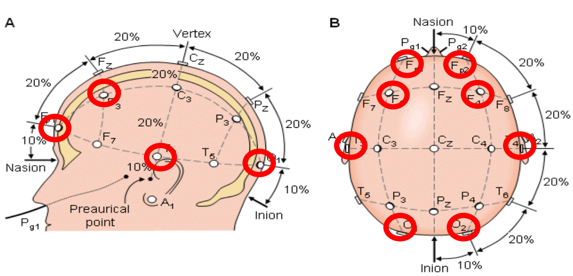


Fig. 1. The 10-20 System of Electrode Placement

The EEG data from 10 healthy volunteers of both genders was then collected. Subjects who participated in the experiment were in good physical condition. We created two databases for use in our experiments, one which included EEG data for facial expressions only, and another which included EEG data for both emotions and facial expressions. The method of recording the two databases is explained in the next section.

2.1 Facial Expression Database

Fig. 2 is an illustration of emotional expressions. This figure was used for recording EEG data about facial expressions (DB1). To define each facial expression, subjects imitated the illustrations in Figure 2. EEG data about expressions was collected at that time.

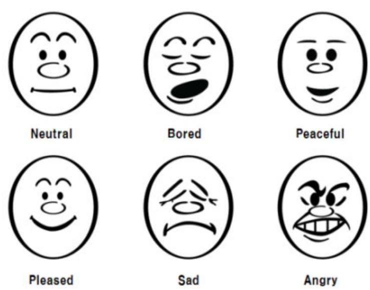


Fig. 2. Illustrations of facial expressions used in collecting EEG data

Before data acquisition, volunteers held a neutral emotional state. While keeping that neutral emotional state, they imitated the facial expressions in the illustrations for a short time, one by one. They only imitated each emotional expression for five seconds in order to avoid adopting that emotion. The timeline used to measure expressions for the expression database is illustrated in Fig. 3. Each expression was held for five seconds. The data set consists of six expressions recorded by each subject. Each subject recorded four different sets of data. The data is used to define the effect of facial expressions on EEG data.

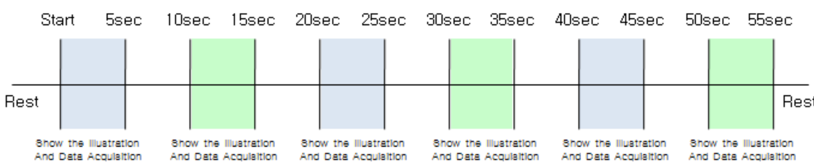


Fig. 3. Timeline used to record the EEGs of facial expressions for the facial expression database

2.2 Emotion and Facial Expression Database

After construction of the expression database, the database of both emotion and facial expression information (DB2) was collected. In order to record this database, we needed to induce emotions in the subjects. The illustrations of expression and pictures of the International Affective Picture System (IAPS) were used to evoke emotions in the emotion and expression database. The feelings of the subjects were excited using the IAPS. IAPS, developed by Lang, is a visual database for emotional research [7].

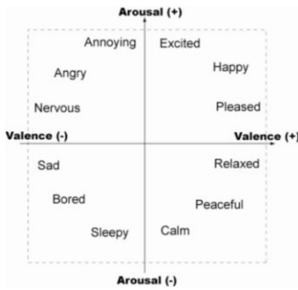


Fig. 4. Thayer's two-dimensional arousal-valence emotional plane

IAPS offers mean and standard deviation of the Arousal, Pleasure and Dominance of each slide. We used these preferences to map emotions to Thayer's two-dimensional arousal-valence emotional plane and group the photos into each emotion. Thayer's emotional plane is presented in Fig. 4, and Fig. 5 shows some examples from the IAPS.

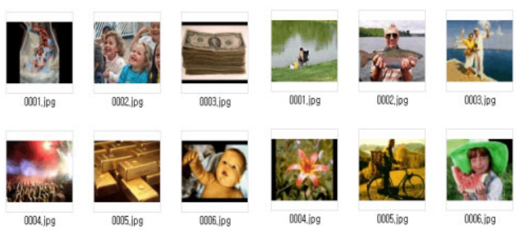


Fig. 5. Example stimulus photos from the IAPS

The procedure for constructing DB2 is as follows. Six IAPS slides for each emotion are presented to subjects for 30 seconds. The subjects look at the slides to arouse their emotions. A facial expression illustration and IAPS are provided to subjects at the same time so that they can imitate the facial expression while feeling the emotion. Data on each emotion was recorded for 30 seconds. Between recordings of each emotion, the subjects were free to relax their emotional state. Data about six different emotions created a set of data. Two sets of data were recorded from each subject. Fig. 6 shows a timeline for the measurement of each emotion.

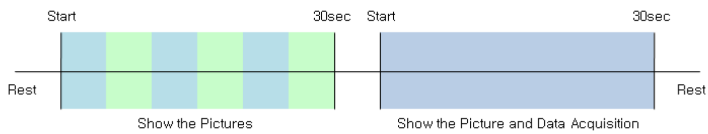


Fig. 6. Timeline of the procedure to record emotions and facial expressions for the database

2.3 Spectrum Subtraction

A spectrum is used frequently in EEG signal processing [3] because the spectrum offers significant information. In this paper, we used a spectrum to represent information about emotions and facial expressions. We subtracted the facial expression spectrum from the emotion and facial expression spectrum using several steps. First, average spectra were calculated from the facial expression database and used to define the mean of an expression in an EEG. Using this mean spectrum, we created an emotional database by subtracting the mean facial expression spectrum from the full spectrum of each frame in the emotion and facial expression database. The subtraction resulted in pure information about emotions in the time dimension. A database of the reconstructed emotional spectrum was used for the recognition experiment to compare results of the original database with results of the reconstructed database. Through the comparison, we can evaluate our proposed system. Fig. 7 presents a flowchart of our proposed system.

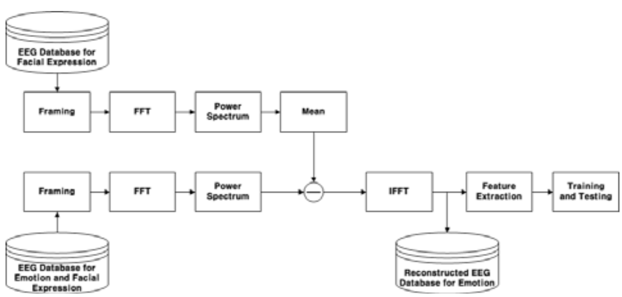


Fig. 7. Flowchart of Proposed System

3 Experiments and Results

The experiments were performed twice, once with the original emotion and facial expression database and again with the reconstructed database. Experimental data was collected from ten subjects (S1~S10). Each volunteer's EEG signal was measured twice, once for the learning database and again for the test database. All of the experiments for emotional models were performed separately and classified by individual emotion: neutral, bored, peaceful, pleased, sad, and angry.

3.1 Feature Extraction

The proposed system used statistical features to define emotional feature vectors. The statistical feature vectors were proposed only for physiological signals. These feature vectors were expanded by fusing the features of physiological signals with those derived from the respective EEG recordings.

The statistical features used to form the proposed feature vectors are defined as follows:

1) The means of the raw signals:

$$\mu_X = \frac{1}{N}\sum_{n=1}^{N} X_n \ . \tag{1}$$

2) The standard deviations of the raw signals:

$$\sigma_X = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(X_n - \mu_X)^2} \ . \tag{2}$$

3) The means of the absolute values of the first differences of the raw signals:

$$\delta_X = \frac{1}{N-1}\sum_{n=1}^{N-1}|X_{n+1} - X_n| \ . \tag{3}$$

4) The means of the absolute values of the first differences of the normalized signals:

$$\bar{\delta}_X = \frac{1}{N-1}\sum_{n=1}^{N-1}|\bar{X}_{n+1} - \bar{X}_n| = \frac{\delta_X}{\sigma_X} \ . \tag{4}$$

5) The means of the absolute values of the second differences of the raw signals:

$$\gamma_X = \frac{1}{N-2}\sum_{n=1}^{N-2}|X_{n+2} - X_n| \ . \tag{5}$$

6) The means of the absolute values of the second differences of the normalized signals:

$$\bar{\gamma}_X = \frac{1}{N-2}\sum_{n=1}^{N-2}|\bar{X}_{n+2} - \bar{X}_n| = \frac{\gamma_X}{\sigma_X} \ . \tag{6}$$

The corresponding feature vector (FVs) is defined as

$$\mathrm{FVs} = \begin{bmatrix} \mu_{X_{ch1}}, \sigma_{X_{ch1}}, \delta_{X_{ch1}}, \bar{\delta}_{X_{ch1}}, \gamma_{X_{ch1}}, \bar{\gamma}_{X_{ch1}}, \dots, \\ \mu_{X_{ch8}}, \sigma_{X_{ch8}}, \delta_{X_{ch8}}, \bar{\delta}_{X_{ch8}}, \gamma_{X_{ch8}}, \bar{\gamma}_{X_{ch8}} \end{bmatrix} . \tag{7}$$

3.2 Recognition Experiments

Recognition experiments were performed using the emotion and facial expression database and the reconstructed database. The test database and trained database were organized for each subject. Likewise, the recognition results were also calculated for each subject. Results of the original and reconstructed data were compared to evaluate the proposed system. A feature vector extracted from EEG was recognized using the k-NN algorithm. The k-NN algorithm classifies the class using the distance between the emotional model and the input data. Recognition rates of k-NN depend on the distance measuring method and the k value. In this paper, test data was classified by the k-NN algorithm when k=5 and the Euclidean method was used for the distance measuring method.

3.3 Experiment Results

Table 1 presents the classification results of the database which has information of emotions and facial expressions. As shown in Table 1, the recognition rate of the original database was at most 29.9%.

Table 1. Emotion recognition results using the database for emotions and facial expressions

(%)	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	Ave.
Angry	23.7	32.2	22	13.6	15.3	61	33.9	16.9	69.5	39	32.7
Neutral	27.1	11.9	57.6	10.2	22	35.6	23.7	11.9	18.6	33.9	25.3
Pleased	37.3	15.3	42.4	10.2	28.8	8.5	39	22	64.4	16.9	28.5
Sad	6.8	6.8	6.8	1.7	27.1	18.6	39	33.9	23.7	28.8	19.3
Bored	22	66.1	40.7	35.6	13.6	27.1	40.7	23.7	27.1	11.9	30.8
Peaceful	37.3	32.2	47.5	54.2	61	52.5	64.4	16.9	20.3	44.1	43.1
Average	25.7	27.4	36.2	20.9	28	33.9	40.1	20.9	37.3	29.1	29.9

Table 2. Emotion recognition results using the reconstructed database

(%)	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	Ave.
Angry	100	89.8	98.3	93.2	33.9	94.9	91.5	76.3	96.6	89.8	86.4
Neutral	79.7	83.1	86.4	72.9	64.4	98.3	76.3	18.6	100	64.4	74.4
Pleased	39	86.4	89.8	81.4	37.3	96.6	84.7	76.3	93.2	83.1	76.8
Sad	62.7	100	100	98.3	83.1	84.7	98.3	78	100	96.6	90.2
Bored	89.8	91.5	84.7	88.1	50.8	93.2	83.1	40.7	69.5	72.9	76.4
Peaceful	89.8	98.3	100	96.6	83.1	96.6	81.4	25.4	100	88.1	85.9
Average	76.8	91.5	93.2	88.4	58.5	94.1	85.9	52.5	93.2	82.5	81.7

The recognition rate of the database reconstructed through spectrum subtraction is presented in Table 2. When using the reconstructed database we obtain a mean accuracy of 81.7% with a maximum of 94.1%. As a result, we confirm that recognition rates increased about 51.8% when compared to the original data. And by comparing Table 1 and Table 2, we can confirm that facial expression has a negative impact on emotion recognition.

4 Conclusions

In this paper, an EEG-based emotion database was reconstructed using spectral subtraction and emotional recognition performance was evaluated. We constructed two databases, one containing EEG readings of emotions and the other containing EGG readings of emotions and facial expressions. We created a third database containing pure information about emotions which was recorded by spectral subtraction, and compared results when using the original facial expression and emotion database and the reconstructed database. As a result, we confirmed that recognition rates increased about 51.8% when comparing the original and reconstructed data. This result shows that facial expression has a negative impact in EEG-based emotion recognition systems. The next stage of this study is to explore practical applications of this data in a real-time system.

Acknowledgments. This research was supported by the MKE(The Ministry of Knowledge Economy), Korea, under the ITRC(Information Technology Research Center) support program supervised by the NIPA(National IT Industry Promotion Agency) (NIPA-2010-(C1090-1021-0008))(NTIS-2010-(1415109527)) and Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(2010-0021411)".

References

1. Schaaff, K., Schultz, T.: Towards Emotion Recognition from Electroencephalographic Signals. In: 3rd International Conference on Affective Computing & Intelligent Interaction, ACII (2009)
2. Takahashi, K.: Remarks on emotion recognition from bio-potential signals. In: Proc. of 2nd Int. Conf. Autonomous Robots Agents, pp. 186–191 (2004)
3. Ko, K.-E., Yang, H.-C., Sim, K.-B.: Emotion Recognition using EEG Signals with Relative Power Values and Bayesian Network. International Jounal of Control, Automation, and Systems, 865–870 (2009)
4. Gao, J., Lin, P., Yang, Y., Wang, P.: Automatic removal of eye-blink artifacts based on ICA and peak detection algorithm. In: 2nd International Asia Conference on Informatics in Control, Automation and Robotics (CAR), pp. 22–27 (2010)
5. Yoo, K.-S., Basa, T., Lee, W.-H.: Removal of Eye Blink Artifacts From EEG Signals Based on Cross-Correlation. In: Proceedings of the, International Conference on Convergence Information Technology, ICCIT 2007, pp. 2005–2014 (2007)
6. Chakraborty, A., Bhowmik, P., Das, S., Halder, A., Konar, A., Nagar, A.K.: Correlation between stimulated emotion extracted from EEG and its manifestation on facial expression. In: IEEE International Conference on Systems, Man and Cybernetics, SMC 2009, pp. 3132–3137 (2009)
7. Lang, P.J.: International Affective Picture System(IAPS): Technical Manual and affective ratings. NIMH center for the Study of Emotion and Attention, Gainsville

Improving Human-Machine Interaction – A Non Invasive Approach to Detect Emotions in Car Drivers

Michael Oehl, Felix W. Siebert, Tessa-Karina Tews,
Rainer Höger, and Hans-Rüdiger Pfister

Leuphana University of Lüneburg, Institute of Experimental Industrial Psychology,
Wilschenbrucher Weg 84a, 21335 Lüneburg, Germany
{oehl,fsiebert,tews,hoeger,pfister}@uni.leuphana.de

Abstract. As cars become increasingly computerized, automatic emotion detection and affective computing provides a promising basis for future-oriented human-computer interaction (HCI) in cars. However, we are still facing severe problems when trying to detect the users' emotional state reliably. This experimental study investigated grip-strength as a new non-invasive method to detect emotions directly in an automobile context. A positive emotion (happiness) and a negative emotion (anger) were examined regarding their influence on grip-strength applied to the steering wheel. Results confirmed and extended preliminary findings: Drivers' grip-strength slightly increased while driving a car when happiness was experienced and especially decreased when anger was experienced. Implications for further research as well as for praxis are outlined.

Keywords: Emotion detection in cars, human-computer interaction, human-centered design, affective computing.

1 Introduction

Human factors are the main reasons of traffic accidents probably not only in the European Union [1]. To counter this, politics focused on improving safety measures that are implemented in the traffic surroundings, e.g., guard railing, while car manufacturers improved passenger protection. Albeit these improvements, research still suggests that emotions are a main cause for maladjusted driving that leads to accidents [e.g., 2, 3, 4]. Therefore, emotions are an important object of applied human factors research aiming at an optimized human-machine interaction. However, while there has been much progress in the research of the relation between especially negative emotions, e.g., anger, maladjusted driving, and accident risk; little attention has been paid to the detection of drivers' emotions in cars. Therefore, the focus of this experimental study lies on the *detection of emotions in automobile drivers*.

As cars become increasingly computerized, automatic emotion detection and affective computing provides a promising basis for future-oriented human-computer interaction (HCI) in cars. While other driver states, such as sleepiness and fatigue can be reliably detected in cars, researchers and manufacturers are still facing severe problems when trying to detect the users' emotional state reliably. Our current study faces

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 577–585, 2011.

this crucial question especially in an automobile HCI context. So the focus of this experimental study lies on the automatic detection of emotions in car drivers following a promising multimodal approach. Since the methods used to detect emotions needed to be applicable in a car, we have only chosen non-invasive measurements that do not interfere with the drivers' ability to control the car, and are in part already implemented in today's state of the art cars. Earlier studies tried to detect emotions in general with a number of non-invasive measurements, e.g., speech [e.g., 5] and facial expressions [e.g., 6].

1.1 Detecting Driver's Emotions

Taking into account the results of Mesken et al. [3], we focused on methods that can be measured constantly and show enough variance, to allow discrimination between different emotions. In Mesken's study participants' facial expression only corresponded with *happiness*, while *anger* and *anxiety* corresponded with a neutral face. Furthermore, Eyben et al. [7] suggest that speech does not occur constantly and that there are emotional states, in which a driver will not talk at all. Hence these limitations of the variables facial expression and speech, a new measure – *grip-strength* – was applied and evaluated in this current experimental study.

The driver's grip-strength applied to the steering wheel can constantly be measured, and a pilot study [8] showed variance in grip-strength that allowed to discriminate between different emotional states. As additional non-invasive measurements, sitting posture and variation in the accelerator pedal were recorded. The sitting posture of drivers is a promising measure to detect different emotions. Overbeeke, Vink and Cheung [9] suggest that distinguishing emotions through analyzing macro movements of seated humans could be possible. Variation in the position of the accelerator pedal has been linked to high mental workload [10], and could therefore correspond with high emotional arousal. Due to the novelty of the measurement of emotion detection by grip-strength in an automobile context, we will focus on this variable in this paper.

Since there is no data to support the claim that only a certain valence (positive or negative) of emotions leads to maladjusted driving, we have chosen happiness as a prototypical positive emotion, and anger as a prototypical negative emotion to be induced in our experiment. Moreover, these two emotions are experienced very frequently by drivers in traffic [3, 11, 12]. We classified emotions following the approach of Scherer's multi-dimensional emotion-space [13]. Scherer suggests that emotions can be classified by two dimensions, i.e., valence and arousal. Anger and happiness rate high on arousal, while anger ranks negative, and happiness ranks positive on the valence dimension. This allows us to evaluate the grip-strength as a measure to distinguish between different emotional valences.

Earlier studies on grip-strength and emotions showed an increase in grip-strength when comparing a neutral drive with an emotional drive [8, 14]. Furthermore results indicated that gender influences the grip-strength reaction [15]. Compared to male participants female participants' grip-strength increased stronger for happiness, while male participants' showed a higher increase in grip-strength compared to female participants when anger was induced. In the anger condition female participants' grip-strength even decreased compared to the neutral condition. These new findings asked for further research.

So our current study focused on evaluating as well as on extending these preliminary findings with more participants, since the pilot study [8, 15] only had a sample size of N = 21. Furthermore, the method for inducing emotions was changed, to achieve better accuracy in the self-reported emotion questionnaire.

The two central research questions addressed in our current experimental study are:

1) Is it possible to detect emotions by driver's grip-strength applied to the steering wheel as a new additional non-invasive measure to detect emotions in an automobile?
2) Is it possible to differentiate emotions that rank different on the dimension of valence but have an equal arousal level?

2 Method

This current experiment was based on a one-factorial design with repeated measurements. The *independent variable* was the *induced emotion in the driver*. The induction was between-subjects three-tiered in *neutral vs. anger* vs. *happiness*. The *dependent variable* was drivers' *grip-strength applied to the steering wheel*.

In the first part of this paragraph, we will explain the procedure of our study. After this, the method to induce emotions is presented in further detail. At the end of this paragraph, the measurement of grip-strength is outlined.

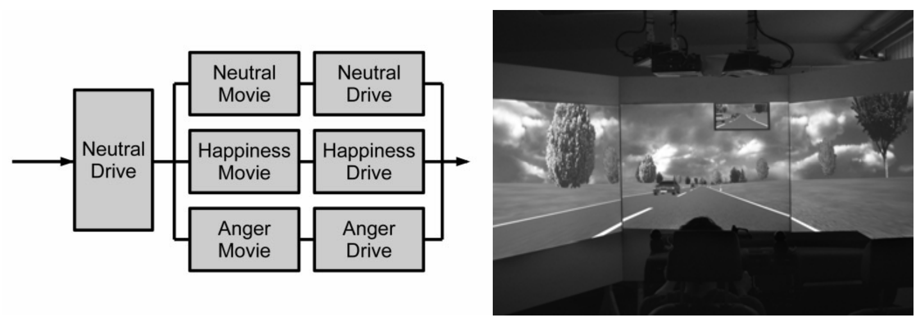


Fig. 1. Experimental procedure and setting in the driving simulator

Participants (N = 59; 40 female) at the age of M = 23.39 years (SD = 4.51) were asked to drive predefined routes in a driving-simulator STISIM DRIVE^{TM} 2.0 (Fig. 1). The routes consisted of country roads, city roads as well as highways and there was sparse oncoming traffic. The design of the routes was modeled after real life traffic-routes, to ensure high validity. Every participant drove a practice route to get used to the driving simulator. After this practice route, the two experimental routes followed in repeated measurements for the dependent variable. The first experimental road was without induced emotions, i.e., the *neutral route*. The second experimental route was with induced emotions, i.e., the *emotional route* (Fig. 1).

After participants finished the neutral route they were shown a film clip to induce a specific target emotion. The method of induction is explained later in this paragraph. Following this, participants drove the second experimental route, i.e., the emotional route. The first neutral and the second emotional experimental route were identical, but this time the surroundings of the road and the appearance of the cars on the road were changed in order to leave participants unaware that they were driving on the same route as before.

In this experiment we induced emotions as the independent variable. Participants were randomly assigned to three different experimental groups, labeled according to the induced emotions *anger*, *happiness,* and *neutral*. The neutral group without induced emotion served as a baseline for the dependent variable grip-strength as well as for the control of sequence effects due to repeated measurements. Anger and happiness were chosen due to their high occurrence in traffic [3, 11, 12]. Emotions were induced through films in this experiment, an efficient method to elicit defined target emotions [16]. We used up-to-date evaluated clips out of film sets which were specific for and ranked highest on the intended target emotion. The stimulus material, i.e., the film clips, was as follows: 'Hannah and her Sisters' (neutral) and 'When Harry met Sally' (happiness / amusement) [17] as well as 'Schindler's List' (anger) [18]. We evaluated the induction of emotions in pre-tests. Additionally, we applied the Self-Assessment Manikin (SAM) as a non-verbal pictorial assessment technique to measure participants' affective reactions [19, 20].

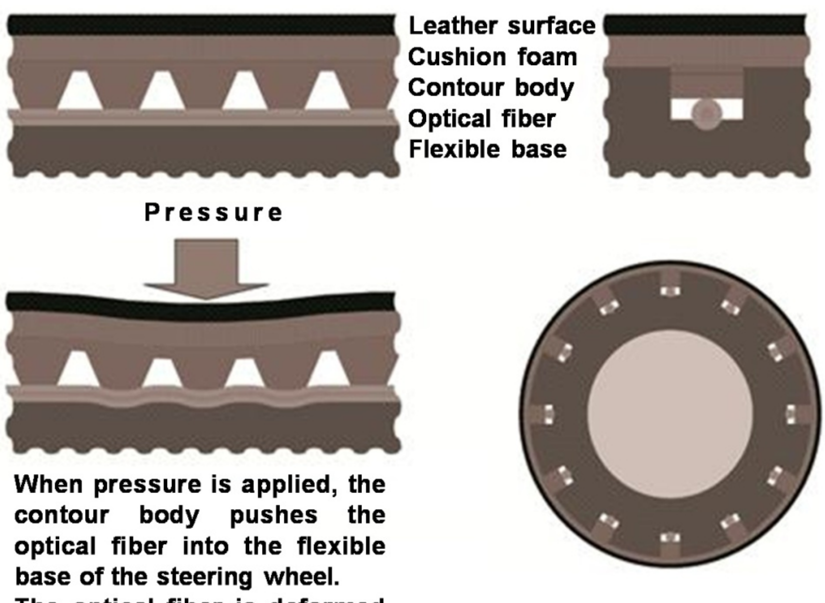


Fig. 2. Schematic structure of the grip-strength sensitive steering wheel (KOSTAL Group)

As *dependent variable* served the *driver's grip-strength* applied to the steering wheel as a new measure to detect driver's emotions in an automobile environment. To measure grip-strength in this experiment, we used a unique steering wheel provided by the KOSTAL Group. The steering wheel had a built in fiber that measured deformation caused by force applied on the steering wheel by participants' hands. Forces were measured all around the steering wheel, giving a unit-less indication for the applied grip-strength. Figure 2 shows the schematic structure of the steering wheel. On the upper left a cross section of the steering wheel is shown while no pressure is applied. On the bottom left the same structure is shown with applied pressure. The upper right shows a single fiber that is used to measure the grip-strength. On the bottom right, the alignment of multiple fibers is shown within the steering wheel.

3 Results

With regard to the Self-Assessment Manikin (SAM) as a non-verbal pictorial assessment technique to measure participants' affective reactions especially the valence of affective reactions is of interest to distinguish between positive and negative emotional states. For the first drive, i.e., the neutral route, participants across all experimental conditions reported no significant differences for the valence of emotional states ($F_{(2,55)} = 2.06$; $p > .05$). But for the second emotional route after the target emotion was induced through films we observed a significant difference of the reported valence ($F_{(2,54)} = 3.37$; $p < .05$). Compared to the neutral route the ratings for the emotional route in the neutral condition were on a ten-points-scale $M = 6.55$ (+0.00; SD = 1.38), in the happiness condition $M = 6.09$ (+0.25; $SD = 1.13$), and in the anger condition $M = 5.41$ (-0.59; $SD = 6.00$). This indicates by trend an effective emotion induction and a carry-over effect of the target emotion in the films for the emotional routes.

Due to technical dropouts we included $n = 55$ participants in the further analyses of the grip-strength data. Based on the unit-less measurement of the grip-strength in our experimental study and additionally due to the large differences in applied grip-strength between participants, especially the relative change of grip-strength within subjects (neutral vs. emotional route) is of interest (Fig. 3).

Participants in the *neutral condition* ($n = 20$) showed some change in grip-strength, comparing the neutral route to the emotional route. During the neutral route, the average grip-strength was $M = 1{,}133.96$ ($SD = 760.60$) while during the emotional route it was $M = 1{,}195.44$ ($SD = 855.84$). This is an increase of $M = 61.48$ or 4.25%. To find possible effects of sequence due to repeated measurements, it is additionally important to analyze the grip-strength of the neutral group (neutral vs. emotional route) in this regard. If there would have been signs of fatigue for example, they would have shown in the data as a decrease in grip-strength. Results suggest that there is no effect of sequence, since there is a small but non-significant increase in grip-strength in the neutral group ($t_{(19)} = -1.25$; $p > .05$).

The *happiness condition* ($n = 18$) showed a small but non-significant increase in grip-strength ($t_{(17)} = -0.49$; $p > .05$) as well. In the neutral route the average grip-strength was $M = 889.71$ ($SD = 674.61$) and in the emotional route it was M = 905.76 ($SD = 666.01$). This is an increase of approximately $M = 16.04$ or 2.16% in grip-strength.

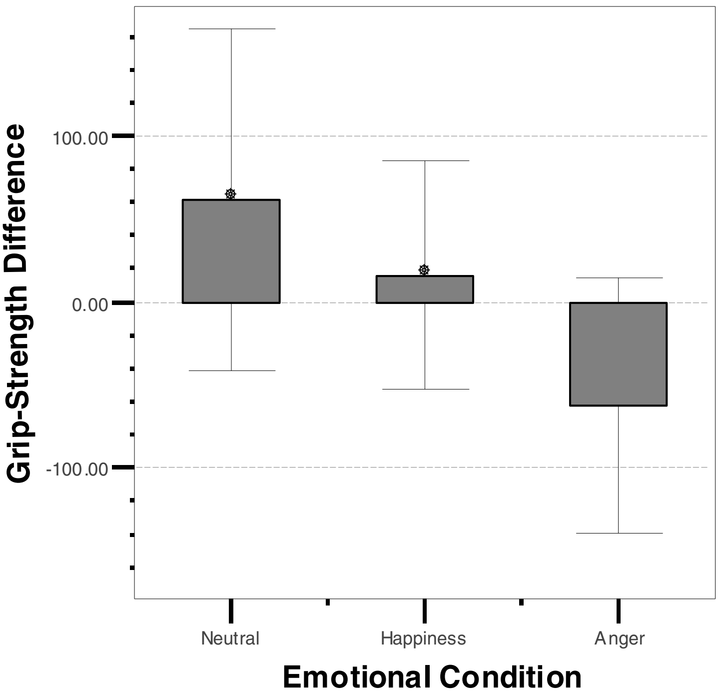


Fig. 3. Mean differences and confidence intervals in grip-strength (neutral vs. emotional route) for the three experimental emotional conditions (neutral, happiness, and anger)

In the *anger condition* (n = 17) there was a nearly significant decrease in average grip-strength within the participants ($t_{(16)} = 1.73$; $p = .05$). During the neutral route their average grip-strength was M = 1,024.12 (SD = 582.49) and during the emotional route they applied a grip-strength of M = 961.32 (SD = 593.36). This equals a decrease in grip-strength of M = -62.80 or approximately -5.82%. These results are shown in figure 3.

A comparison of the changes in grip-strength (neutral vs. emotional route) across the experimental conditions just failed to be significant ($F_{(2,52)} = 2.33$; $p = .11$). Nevertheless, a direct comparison of the grip-strength changes between the neutral and the anger condition yielded a significant difference ($t_{(35)} = 1.97$; $p < .05$).

Further analysis of the data suggested that there were no significant differences in the applied grip-strength between men and women in this experiment for the experimental conditions ($F_{(1,49)} = 0.59$; $p > .05$) nor any significant interaction of gender and experimental condition ($F_{(2,49)} = 0.23$; $p > .05$).

4 Conclusions

So far little attention has been paid to the applied automatic detection of drivers' emotions in cars. Researchers and manufacturers are still facing severe problems when trying to detect the users' emotional state reliably. Therefore, this experimental study focused on the automatic detection of emotions in automobile drivers especially with non-invasive measurements applicable in a car and which do not interfere with the drivers' ability to control the car.

Due to the novelty of the measurement of emotion detection by grip-strength in an automobile context, we focused on this variable in the current paper. Previous studies on grip-strength and emotions showed an increase in grip-strength when comparing a neutral drive with an emotional drive [8, 14]. Furthermore, results indicated that gender influences the grip-strength reaction [15]. The results of our current experimental study contrasted and extended these preliminary findings.

Our first research question asked if it was possible to detect emotions by drivers' grip-strength applied to the steering wheel as a new additional non-invasive measure to detect drivers' emotions in an automobile. Our results suggest that grip-strength can contribute to the detection of emotion. This has to be discussed with respect to our second research question, i.e., if it was possible to differentiate emotions that rank different on the dimension of valence but that have equal arousal levels (happiness vs. anger). Grip-strength measures were slightly but not significantly increased for the happiness group. We observed a comparable increase as well for the neutral condition. However, we found a significant decrease in grip-strength for the anger condition. Since we have neither found any gender differences in grip-strength across the three experimental emotional conditions in the current experimental study nor any differences in the general grip-strength applied to the steering wheel in contrast to previous research [15, 21], our current results with a larger sample are consistent by trend with the female participants' results in our pilot study [15].

These findings suggest that grip-strength can help to differentiate between a non-emotional state of the driver and a critical emotional state like anger. Since critical emotional states lead to maladjusted driving, this differentiation could help in preventing accidents. Nevertheless, we are still facing open research questions, i.e., why is grip-strength deceasing for anger and not for happiness albeit these two emotional states are both associated with high physiological arousal? In this respect a video-based motion tracking of the driver and thus an activity analysis could be insightful [22]. Drivers experiencing anger may have a higher variability of their hand movements and might not stick constantly to the steering wheel. Additionally, an evaluation of drivers' grip-strength behavior with low-arousal emotions of different valence as well as in real life scenarios is needed.

For future research it is also important to refine the measurement of grip-strength. The steering-wheel provided to us, was a unique prototype. Therefore fine-tuning in the algorithm of the calculation of grip-strength may be possible.

Taking into account the results of this study, additional non-invasive measurements to detect drivers' emotions should be considered in order to frame an applied human-centered HCI approach of emotion detection in cars. For example facial expressions have shown to be quite useful for the detection of drivers' happiness but rather difficult for the detection of anger [3]. Besides, other promising new measures like sitting posture and variation in the accelerator pedal should be considered as well.

Our current results showed that grip-strength can contribute to a more valid detection of emotions of car drivers within a multimodal approach. In its application in cars this research could lead to a more human-centered design of HCI in terms of affective computing, e.g., adaptive advanced driver assistance systems (ADAS) depending on the emotional state of the driver and thereby increasing safety to traffic as well as increasing the driver's comfort by regulating his emotional state. Also the implementation of our method of detection would be simple and feasible, since the bigger part of the necessary technology is already available in state of the art automobiles. The ADAS may be realized within a multi-methodic system, taking into account additional non-invasive measures, e.g., facial expressions or speech, thereby increasing the validity of the emotion detection.

Acknowledgments. This research was funded by grant FS-Nr. 2006.63 from the 'Arbeitsgruppe Innovative Projekte' (AGIP) of the Ministry of Science and Culture, Lower Saxony, Germany. We would like to thank the Volkswagen Group (Group Research Electronics and Vehicle Research / HMI and Human Factors) as well as the KOSTAL Group for technical support. Many thanks go to Patricia Brunkow, Johanna von Stebut, Raphael Seitz, Christoph Ruth and Marco Wiethof for their help in conducting the experiment as well as to Heiko Witthöft for his help in analyzing the experiment and his support in applied computer science.

References

1 European Commission: White paper-European transport policy for 2010: time to decide. Office for Official Publications of the Europe Communities, Luxemburg (2010)

2 Deffenbacher, J.L., Oetting, E.R., Lynch, R.S.: Development of a driving anger scale. Psychological Reports 74, 83–91 (1994)

3 Mesken, J., Hagenzieker, M.P., Rothengatter, T., De Waard, D.: Frequency, determinants, and consequences of different drivers' emotions: An on-the-road study using self-reports (observed) behaviour, and physiology. Transportation Research Part F 10, 458–475 (2007)

4 Nesbit, S.M., Conger, J.C., Conger, A.J.: A quantitative review of the relationship between anger and aggressive driving. Aggression and Violent Behavior 12(2), 156–176 (2007)

5 Scherer, K.R., Johnstone, T., Klasmeyer, G.: Vocal expression of emotion. In: Davidson, R.J., Goldsmith, H., Scherer, K.R. (eds.) Handbook of Affective Sciences, pp. 433–456. Oxford University Press, Oxford (2003)

6 Cohn, J.F., Ekman, P.: Measuring facial action. In: Harrigan, J.A., Rosenthal, R., Scherer, K.R. (eds.) The New Handbook of Methods in Nonverbal Behavior Research, pp. 9–64. Oxford University Press, New York (2005)

7 Eyben, F., Wöllmer, M., Poitschke, T., Schuller, B., Blaschke, C., Färber, B., Nguyen-Thien, N.: Emotion on the Road - Necessity, Acceptance, and Feasibility of Affective Computing in the Car. Advances in Human Computer Interaction (AHCI), Special Issue on Emotion-Aware Natural Interaction 2010, 17 (2010), `http://www.hindawi.com/journals/ahci/2010/263593.html`

8 Siebert, F.W., Oehl, M., Pfister, H.-R.: The Measurement of Grip-Strength in Automobiles: A New Approach to Detect Driver's Emotions. In: Karwowski, W., Salvendy, G. (eds.) Advances in Human Factors, Ergonomics, and Safety in Manufacturing and Service Industries, pp. 775–783. CRC Press, Boca Raton (2010)

9 Overbeeke, K.J., Vink, P., Cheung, F.K.: The emotion-aware office chair. In: Helander, Khalid, Tam (eds.) Proceedings of the Int. Conference on Affective Human Factors Design, pp. 262–267. Asean Academic Press, London (2001)

10 Rakauskas, M.E., Gugerty, L.J., Ward, N.J.: Effects of naturalistic cell phone conversations on driving performance. Journal of Safety Research 35, 453–464 (2004)
11 Levelt, P.B.M.: Praktijkstudie naar emoties in het verkeer [Emotions in Traffic]. SWOV Report R-2003-08. SWOV, Leidschendam, Netherlands (2003)
12 Oehl, M., Roidl, E., Frehse, B., Suhr, J., Siebert, F.W., Pfister, H.-R., Höger, R.: Das Emotionsspektrum von Autofahrern im Straßenverkehr [The emotional spectrum of car drivers in traffic]. In: Petermann, F., Koglin, U. (eds.) 47 Kongress der Deutschen Gesellschaft für Psychologie, Abstracts, pp. 371–372. Pabst Science Publishers, Lengerich (2010)
13 Scherer, K.R.: What are emotions? And how can they be measured? Social Science Information 44(4), 695–729 (2005)
14 LaMont, K.: Effect of Induced Emotional State on Physical Strength. In: Kurzer, N. (ed.) Perspectives in Psychology, pp. 3–9. University of Pennsylvania, Philadelphia (2000)
15 Siebert, F.W., Oehl, M., Pfister, H.-R.: Gender Differences in Grip-Strength Depending on Arousal and Valence of Emotions. In: Haack, J., Wiese, H., Abraham, A., Chiarcos, C. (eds.) Proceedings of KogWis 2010 – 10th Biannual Meeting of the German Society for Cognitive Science, p. 227. Universitätsverlag Potsdam, Potsdam (2010)
16 Rottenberg, J., Ray, R.D., Gross, J.J.: Emotion Elicitation Using Films. In: Coan, J.A., Allen, J.J.B. (eds.) Handbook of Emotion Elicitation and Assessment, pp. 9–28. Oxford University Press, Oxford (2007)
17 Hewig, J., Hagemann, D., Seifert, J., Gollwitzer, M., Naumann, E., Bartussek, D.: A revised film set for the induction of basic emotions. Cognition & Emotion 19(7), 1095–1109 (2005)
18 Schaefer, A., Nils, F., Sanchez, X., Philippot, P.: Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers. Cognition & Emotion 24(7), 1153–1172 (2010)
19 Lang, P.J.: Behavioral treatment and bio-behavioral assessment: computer applications. In: Sidowski, J.B., Johnson, J.H., Williams, T.A. (eds.) Technology in Mental Health Care Delivery Systems, pp. 119–137. Ablex, Norwood (1980)
20 Bradley, M.M., Lang, P.J.: Measuring Emotion: The Self-Assessment Manikin and the Semantic Differential. J. Behav. Ther. & Exp. Psychiat. 25(1), 49–59 (1994)
21 Eksioglu, M., Kizilaslan, K.: Steering-wheel grip force characteristics of drivers as a function of gender, speed, and road condition. International Journal of Industrial Ergonomics 38, 354–361 (2008)
22 Cheng, S.Y., Trivedi, S.P., M.M.: Multi-spectral and multi-perspective video arrays for driver body tracking and activity analysis. Computer Vision and Image Understanding 106, 245–257 (2007)

Emotion Recognition Using Biological Signal in Intelligent Space

Kanlaya Rattanyu[1] and Makoto Mizukawa[2]

[1] Graduate School of Functional Control Systems Engineering, Shibaura Institute of Technology, Tokyo, Japan
[2] Department of Electrical Engineering, Shibaura Institute of Technology, Shibaura Institute of Technology, Tokyo, Japan
{m709502,Mizukawa}@sic.shibaura-it.ac.jp

Abstract. In this study, we focus on emotion recognition for service robots in the living space based on Electrocardiogram (ECG). An emotional state is important information that allows a robot system to provide appropriate services in way that are more in tune with users' needs and preferences. Moreover, the users' emotional state can be feedbacks to evaluate user's level of satisfaction in the services. We apply a diagnosis method that uses both inter-beat and within-beat features of ECG. The post hoc tests in Analysis of Variance (ANOVA) showed that our approach satisfies more confidence level of difference between emotions than conventional methods. Our system design was based on wireless and wearable biological sensor for mobility and convenience of users' daily lifestyle.

Keywords: Emotion Recognition, ECG, Intelligent Space.

1 Introduction

In order to provide a suitable service to a user, computers should be more intelligent and more like human beings. Also, they should be able to understand what users feel.

In emotion recognition, we focus on a biological approach that has advantages over other approaches because people's emotions vary in different ways according to environment, culture, education and so on. However, people's emotions are very similar in terms of biological signals. People can hide their emotions from outside appearances but they cannot hide their emotions in biological signals. Biological sensors can solve a limitation of the visual technique that requires a frontal view and clear face to detect expressions. In addition many researchers tried to decrease size of biological sensors for using in health care monitoring. This improvement makes the sensors suitable for our system because it is small enough to be use without any interference to daily life.

Different emotional expressions produce different changes in ANS activity [1]. Paul Ekman et.al found in their experiments that heart rate increased more in anger and fear than in happiness. The main limitation of emotion recognition by using only ECG signals is it can categorize emotion into a few categories such as positive/

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 586–592, 2011.

negative feeling [2,3,4], feeling of being stressed/relaxed [5,6], or fear/ neutrality [7]. Some studies (e.g., [8,9,10,11]) overcome this limitation by combined ECG with other physiological signals that are related with organs that affected by the ANS as showed in table 1. Among these studies, some correlates between emotion and ECG could be identified: increase of heart rate associated with fear (e.g. [9]) and anger (e.g., [8]), increase of heart rate variability associated with stress (e.g. [6]). However some results were controversial: sadness has been found to sometimes lead to an increase (e.g., [12]) and sometimes lead to a decrease (e.g., [9]) of heart rate.

These previous studies extracted only inter-beat information of ECG such as RR-interval or heart rate (HR) time series. Some researches use statistical data of this information (min, max, average, standard deviation of normal-to-normal of R-R intervals (SDNN), Root Mean Square of Successive Different of RR intervals (RMSSD), and heart rate variability (HRV)) in order to maximize efficiency of ECG. We proposed to use ECG's inter-beat features together with within-beat features in our recognition system which we will be describing detail in the next section.

2 Methodology

2.1 System Structured Environment in Intelligent Space

In this system we observed a person in the wide area living space as shown in figure 1. A user wore the RF-ECG that collected data from the movement of the heart's index and sent it to a base station that was connected to a personal computer (PC). This system was designed to be mobile that provided the user more freedom to move while the system monitored the person continuously.

2.2 Wireless ECG Sensor (RF-ECG)

The sensor is selected based on three important criteria. The first criterion was that its signal had to be strongly related with the human emotion. The second criterion was that the sensor had to adhere to human skin without discomfort. The last criterion was

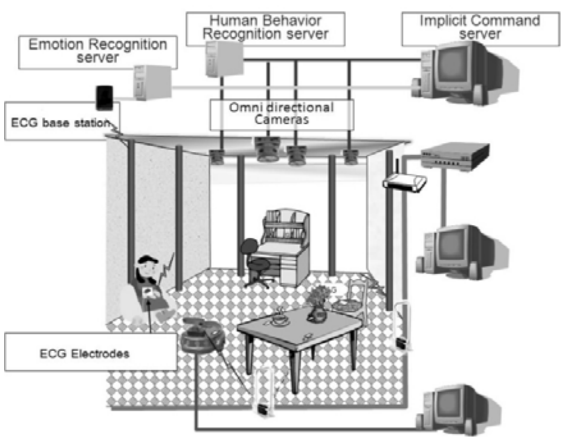


Fig. 1. Structured Environment in Intelligent Space

that the sensor has to be wearable and convenient for use in normal daily life. In this study, the RF-ECG biosensor kit that enables wireless medical monitoring was selected as shown in figure 2. This sensor can record and wirelessly transmit ECG signals to the server with 204 Hz. The sensor utilizes low power consumption RF transmission, which purportedly enables it to broadcast a constant signal for up to 48 hours on a single charge. The sensor used in this study was a low weight (12 g) and small size sensor (40 mm x 35 mm x 7.2 mm). The wireless RF transmitter had an open area range of up to 15 m as shown in table 1.

Table 1. Specification of the RF-ECG monitoring devices

Features	***Specification***
Size and shape	40mm x 35mm x7.2 mm
Material	ABS resin (plastic)
Wireless type	Advanced 2.4 G and Low Band Power Data Comunication
Transmit power	1 mW (0dbm)
Transfer rate	1 Mbps
Protocal	CRC with a proprietary
Current consumption	2-2.4 mA
Distance	Approximately 15 m
Up time	48 hours
Low-frequency cutoff	0.05 Hz - 0.32 Hz
Sampling rate	204 Hz, 102 Hz
High-frequency cutoff	100 Hz

2.3 Emotion Recognition

ECG signal was sampled with a sampling frequency of 204 Hz. Then digital signal was transmitted wirelessly to the server as shown in figure 2.

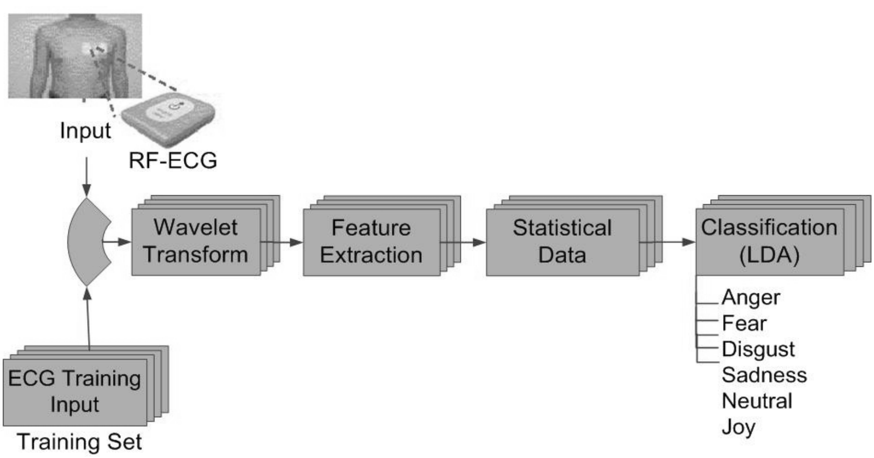


Fig. 2. Emotion recognition by using RF-ECG sensor

Annotation of the ECG (Wavelet Transform). The continuous wavelet transforms (CWT) and fast wavelet transforms (FWT) were used for automatic annotation of the ECG cardio cycle[old8]. The annotation method consists of two phases: QRS detection followed by P, T wave location.

- QRS detection: To amplify the QRS complex and separate low frequency (P and T waves) and high frequency (noise), the CWT transform is applied at 12Hz with an inverse wavelet. The CWT spectrum obtained is further filtered with FWT using an interpolation filter to remove frequency content below 30Hz and the rest of the spectrum is denoised with a hard threshold using a MINIMAX estimate. The reconstructed ECG signal after denoising contains only spikes with nonzero values at the location of the QRS complexes.
- P, T waves detection: After QRS complexes are detected the intervals between them are processed for detection of P and T waves.

Feature extraction. After we located the location of each wave on ECGs, several parameters that indicate each part of heart's activity were calculated. In this process, we calculate not only inter-beat information of ECG (RR-interval or HR) but also within-beat information of ECG (PR, QRS, ST, QT intervals, PR, and ST segments).

Statistical data. To prevent data loss and reduce the number of the features, we used the statistic data (mean, and standard deviation) of each parameter to be the features. Note: there are two types of standard deviation between heart beats (SDNN and RMSSD).

Different people have different biological signals. To avoid this effect, each feature (HR, SDNN, RMSSD, QT, STDEV(QT), PR, STDEV(PR), STDEV(QRS), ST, STDEV(ST)) was normalized by subtracting each parameters from its mean in the neutral emotion.

Classification. Linear Discriminant Analysis (LDA), and Adaptable K-Nearest Neighbor(A-KNN)[13] are used to classify emotion into six categories(anger, fear, disgust, sadness, neutral, and joy). To evaluate the recognition performances, we separate data into two set. There are 147 data for training and 147 data for testing.

3 Experiments

3.1 Picture Database

Our experiments were based on International affective picture system (IAPS) database [old10] that contained 20 sets of 60 pictures. Each picture induced a variety of emotion and has pre-rated of emotion. In experiment, we selected 60 pictures.

3.2 Participants

This study was conducted with 6 subjects, including 5 healthy males (mean±SD age = 27.2±3.63 years) and one female (age =33). We didn't examine many subjects since each subject was recorded with 60 ECG signals. The total number of samples was 360(6x60) samples that was sufficient.

3.3 Data Collection

In the experiments, we applied the same process with IAPS experiments. However, we changed from pencil-paper based to computer based for questionnaire section to reduce the effort for management of questionnaires. We provided the same questionnaires to the all subjects. When the subjects did not feel emotions promptly in IAPS, we did not apply the data to an experiment. In the experiment, we showed 60 pictures to the subjects one by one. Each trial began with a preparation step to train the subjects to familiar with and understand the experiment. We attached ECG electrodes on the subject's chest. The picture to be rated was presented for 6 seconds, and immediately after the picture left the screen, the subject made their ratings. A standard 15 seconds rating period was used, which allowed ample time for the subjects to finish the questionnaire. We started to measure the ECG signal at the same time the first picture was presented, so we were able to separate the ECG signal into 6 second of 60 signals per subject.

4 Results and Discussions

4.1 Post Hoc Tests in Analysis of Variance (ANOVA)

The Least Significant Difference (LSD) test was used to explore all possible pair-wise comparisons of means comprising an emotion factor using the equivalent of multiple t-tests. In this section, we compare two techniques.

- Three Features approach: This is traditional technique that applies inter weaves' information (HR, SDNN, RMSSD) for emotion recognition.
- Eleven Features approach: Our proposed technique that applies both inter-beat and within-beat information (HR, SDNN, RMSSD, QT, STDEV(QT), PR, STDEV(PR), STDEV(QRS), ST, STDEV(ST)).

The Post Hoc Test's results are showed in figures 3 and 4. In each pair of emotions, we selected the highest value of confidence interval among three features approach in figure 3 and eleven features approach in figure 4. As shown in figure 3 and 4, the eleven feature approach had more confidence-level than the traditional method and also the classification accuracy is higher as shown in table 2.

Table 2. The comparison of emotion recognition's accuracy between Three Feature and Eleven Feature approaches

Emotion	*Three Features Approach*	*Eleven Features Approach*
Anger	68.57 %	60.00 %
Fear	61.25 %	32.50 %
Disgust	57.14 %	25.71 %
Sadness	64.38 %	45.21 %
Neutral	60.53 %	31.58 %
Joy	56.79 %	28.40 %
Total	**61.44 %**	**37.23 %**

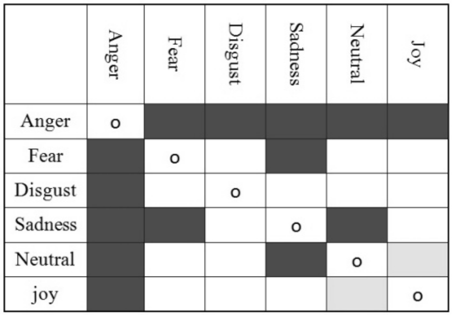


Fig. 3. Three Features Approach Post Hoc Tests in ANOVA

5 Conclusion and Future Work

This system, we focus on emotion recognition in the living space. The RF-ECG sensor is a wearable sensor that uses a wireless connection to the server so the system can monitor the user all the time. Therefore the user is able to move around freely in the living space. We apply a diagnosis method that uses of both inter-beat and within-beat of ECG signals for the emotion recognition with improved accuracy. We reduce the amount of raw data by using analyzed value of ECG signals and statistical data in emotion recognition.

In the future work, we plan to combine more biological signals such as Respiration (RESP), and Skin temperature to improve the accuracy of recognition rate. Physiological signals could be used in health care monitoring in tandem with emotion recognition. While this study deals with very new systems and there are some subjects for research which remain unimplemented such as service generation depends on emotions, they may be useful in the creation of new types of processing systems in the near future.

Acknowledgements. The authors would like to express their gratitude to the professor Michiko Ohkura (Shibaura Institute of Technology) for her useful advice support of the experiments and equipments.

References

1. Paul Ekman, R.W.L., Friesen, W.V.: Automomic Nervous System Activity Distingquishes among Emotions. Science, New Series 221(4616), 1208–1220 (1983)
2. Chun-Han Yang, K.-L.L.Y.-H.K., Wang, J.-L., Cheng, K.-S.: Negative Emotion Detection Using the Heart Rate Recovery and Time for Twelve-Beats Heart Rate Decay After Exercise Stress Test. In: International Joint Conference on Neural Networks (IJCNN), pp. 1–6. IEEE, Los Alamitos (2010)
3. Thayer, J., Siegle, G.: Neurovisceral integration in cardiac and emotional regulation. IEEE Engineering in Medicine and Biology Magazine 21(4), 24–29 (2002)

4. Wu, W., Lee, J.: Improvement of HRV Methodology for Positive/Negative Emotion Assessment. In: 5th International Conference on Collabarative Computing:Networking, Application and Worksharing, Collaborate-Com 2009, pp. 1–6. IEEE, Los Alamitos (2009)
5. Lee, C., Yoo, S.: ECG-based Biofeedback Chair for Selfemotion Management at Home. In: International Conference on Consumer Electronics, ICCE 2008. Digest of Technical Papers, pp. 1–2. IEEE, Los Alamitos (2008)
6. Rodriguez, J.D., Santos, L.: Comparative Analysis Using the 80-Lead Body Surface Map and 12-Lead ECG With Exercise Stress Echocardiograms. Journal of Diagnostic Medical Sonography 22(5), 308–316 (2006)
7. Vrana, B.N.C.S.R., Lang, P.J.: Fear Imagery and text processing. The International Journal of the Society for Psychophysiological Research (Psychophysiology) 23(3), 247–253 (2007)
8. Barbara, C.B., Fredrickson, L., Mancuso, R.A., Tugade, M.M.: The Undoing Effectof Positvie Emotion. Motivation and Emotion 24(4), 237–258 (2000)
9. Christie, I.C.: Multivariate Discrimination of Emotion-Specific Autonomic Nervous System Activity. Master's thesis, Virginia Polytechnic Institute and State University (2002)
10. Lee, C.K.: Using Neural Network to Recognize Human Emotions from Heart Rate Variability and Skin Resistance. In: International Conference of the IEEE-Engineering in Medicine and Biology Society (IEEE-EMBS), pp. 5523–5525. IEEE, Los Alamitos (2005)
11. Kim, J., Andre, E.: Emotion Recognition Based on Physiological Changes in Music Listening. IEEE Transactions on Pattern Analysis and Machine Intelligence 30(12), 2067–2083 (2008)
12. Pedro Montoya, J.J.C., Schandry, R.: See red? Turn pale? Unveiling Emotions Cardiovascular and Hemodynamic Change. Spanish Journal of Psychology 8(001), 79–85 (2005)
13. KanlayaW, L.D., M., M.: Virtual Object for Evaluating Adaptable K-Nearest Neighbor Method Solving Various Conditions of Object Recognition. In: ICROS-SICE International Joint Conference 2009 (ICCAS-SICE 2009), p. 4338. IEEE, Los Alamitos (2009)

Intentionality in Interacting with Companion Systems – An Empirical Approach

Dietmar Rösner[1], Rafael Friesen[1], Mirko Otto[1], Julia Lange[2], Matthias Haase[2], and Jörg Frommer[2]

[1] Department of Knowledge Processing and Language Engineering, Otto-von-Guericke University, Magdeburg, Germany
{roesner,friesen,miotto}@ovgu.de

[2] Department of Psychosomatic Medicine and Psychotherapy, Otto-von-Guericke University, Magdeburg, Germany
{joerg.frommer,matthias.haase,julia.lange}@med.ovgu.de

Abstract. We report about a WOZ experiment with a carefully designed scenario that allows to investigate how users interact with a companion system in a mundane situation with the need for planning, re-planning and strategy change. The data collection from the experiments comprises multimodal records (audio, video, biopsychological parameters) and transcripts of the verbal interaction, and all subjects fill out a battery of well established psychometric questionnaires about various aspects especially of their personality. This will allow to correlate observed behaviour and detected affects and emotions with measured aspects of the personality of subjects and is expected to serve as a basis for defining a typology of users. In addition, a subgroup of the subjects takes part in semiformal in-depth interviews that focus on retrospective reflexion of the users' subjective experience during the experiments and especially on the intentionality that users ascribed to the system during the course of interaction.

Keywords: Intentionality, Wizard of Oz, Companion Systems, Emotion, Multimodal.

1 Introduction

Companion systems are an important class of emerging future devices. These can be defined as: "[. . .] conversationalists or confidants – not robots – but rather computer software agents whose function will be to get to know their owners, who may well be elderly or lonely, and focusing not only on assistance via the internet (contacts, travel, doctors etc.) that many still find hard to use, but also on providing company and companionship, by offering aspects of personalization" [16].

In order to realize such systems, psychological and social issues of human-computer-interaction (HCI) have to be taken into account [16]. However, studies in this field concentrate on the users' emotion, reactance and cooperativeness.

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 593–602, 2011.

One aspect, which is furthermore relevant for the acceptance and usability of such systems, has not been investigated enough: the issue of intentional stance [3] towards companion systems. Hereby, explanations and predictions of the systems' behaviour are based upon the presumption that the system holds certain information, pursues certain goals and chooses a behavioural pattern on that background, which is rational and appropriate to the situation. This means that humans ascribe intentional states to the system such as hopes, worries, perceptions, emotions etc. We have therefore designed and implemented a Wizard of Oz (WOZ) experiment 'last minute' that allows to empirically investigate the following issues: intentional stance and user emotions during HCI, the risk of breaking off the dialogue, and the classification of different user-types. User types are discriminated by e.g. specific markers in spoken language, certain psychological constructs, socio-demographic characteristics and the users' subjective experience of HCI. This experiment is currently conducted with users from two age groups: young adults between 18 and 28 and seniors above 60, balanced in gender and educational level.

In the following we elaborate on the design rationale of the szenario of the WOZ experiment and indicate first findings from the ongoing experiments.

2 The Intentional Stance in Interacting with Companion Systems

The philosopher Daniel C. Dennett described three stances humans can adopt towards a biological (including people) or non-biological system to explain or predict its behaviour: physical, design and intentional stance [3].

By using the term "intentional" Dennett refers to the philosophical term "Intentionality" as described by Franz Clemens Brentano [4,3]. He writes "the intentional strategy consists of treating the object [. . .] as a rational agent with beliefs and desires and other mental stages exhibiting what Brentano and others call intentionality" ([4] p. 15).

Companion systems are designed for reacting individually and empathetically to users. Their construction plan and the physical processes inside the systems are complex in a way that an average user is unable to predict or explain the behaviour of the system in the design or physical stance. Hence, for an average user adopting the intentional stance is the most natural and pragmatic way of predicting and explaining the behaviour of such systems successfully. It is desirable that users assume positive intentions in companions system like helpfulness, trust-worthiness and empathy. These are able to increase the acceptance and usability of companion systems, whereas assumptions of negative intentions like malice, pursuit to dominance and poor willingness to cooperate may contribute to their decrease.

3 A Wizard of Oz Experiment

The WOZ experiment was designed to investigate how users interact with a companion system in a mundane situation with the need for planning, replanning and strategy change. The dialogue processes experimentally induce critical events that provoke negative emotions and as a consequence possibly break off the dialogue. Our aims are (1) to find out features in users' speech content that allow a prediction of dialogue crisis, e.g., manifestations of negative emotions, (2) to investigate the usefulness of an empathic intervention in dialogue crisis and (3) to reconstruct how users experienced the interaction and its critical events, especially which emotions and intentional ascriptions occurred. Critical events are not induced by a "dysfunction" of the companion system but by an enforced strategy-change and additionally occurring constraints that complicate task completion. The software architecture is described in [13]. The WOZ experiment comprises two submodules [6]. In the first submodul "initial narrative" the subject undergoes an initialisation by a personalization dialogue. This simulates that the CS is individualized for one user to adapt to him or her, to stay with him or her for a long time and to support him or her in everyday life tasks - the most relevant characteristics of companion systems. The second submodule "last minute" is described in detail in the following paragraph.

3.1 Wizard of Oz Scenario

The systemic functional linguistics model of Halliday [8] analyzes three metafunctions of language: ideational, interpersonal and textual. The ideational metafunction bears the content of speech. Speech in current HCI is focussed primarily on this metafunction. To be reliable and empathic companion systems have to adapt to the user's situation, skills, preferences, aims, needs and emotions. Thus companion systems have to communicate not only on the ideational metafunction but employ also the textual and interpersonal metafunction.

This is simulated and investigated in the second submodule "last minute" consisting of six stages (Fig. 1).

1. The user is asked to imagine a situation in summer when he as a surprise gets informed to have won a two weeks vacation. The prize includes as well the opportunity to select all necessary items for the travel from an online catalogue organized in a number of categories (e.g. coats, underwear, etc.).

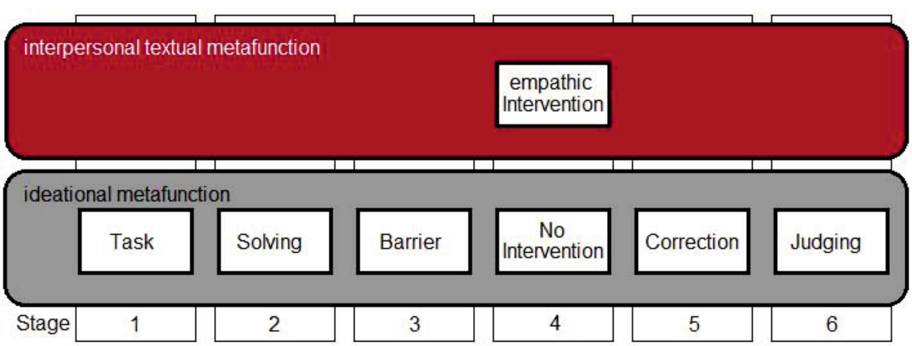


Fig. 1. Stages of the WOZ experiment

For this selection fifteen minutes are available, because then the taxi to the airport will arrive and the packed suitcase will be available at check in. This cover story is provided very detailed to stimulate ego involvement and imagination of the subject. The task to choose equipment and to pack a suitcase is introduced and some (time limit), but not all (weight limit of suitcase) limitations are given.
2. The subject can choose items in a uniform procedure from 12 categories in a predefined sequential order. The system informs the subject when a limitation (weight limit of the suitcase) is infringed and provides information about the task's status.
3. The system informs the subject that the destination is in the southern hemisphere and the forecast estimates temperatures around 0 °C instead of the summer temperatures assumed by the subject. This information lets the subject realize that he had aimed a wrong goal and has now to change his strategy. This might lead to frustration and anger.
4. The system addresses the subject using the interpersonal and textual metafunctions with increased intensity asking questions about the subject's feelings and motivation. Such selected dialogue strategies create the chance for reflection and expression of anger and provide empathic help based on the principles of Rogers' paradigm of client centered psychotherapy [15].
5. The subject can select from chosen categories a second time to readjust the equipment to the cold weather. The system's speech is focussed primarily on the ideational metafunction again.
6. After finishing the selection from the remaining categories, the subject is asked to rate his own performance: how content he is with the packed suitcase and if the subject would start the trip with the suitcase's content.

3.2 An Example Interaction

The following excerpts are anonymized and translated examples of a dialogue in German from our first experiments. Please note that we tried to preserve the subject's talking style. (U=user, S=system, [...] parts left out)

Stage 1: Personalisation (Task)

```
   [...]
S: At the beginning some information about you is necessary. Can you
   please introduce yourself?
U: Yes. My name is Max, coming Magdeburg, and doing this test.
S: The system needs more information to adapt itself to you. Can you
   please make a statement about the following points: Your age, your job,
   your jobs location, your family, your dress size, your size of shoe.
U: Yes, my size of shoe is 42, I am 22 years old, my birthday is
   1.1.1990, I am single, have no girlfriend, work in construction,
   everywhere in Magdeburg, and yes.
S: Please complete the information about your body size.
   [...]
```

```
S: Please describe a recent event that you enjoyed.
   [...]
S: Please describe a recent event that annoyed you.
   [...]
```

Stage 2: Selection (Solving)

```
   [...]
U: Next section.
S: You can now choose from the section 'coats and cloaks'.
U: Yeah. I would take a macintosh, or two.
S: A macintosh was added. You can continue.
U: And a jacket.
S: A jacket was added. You can continue.
   [...]
```

“Weight Limit"

```
U: Two swimming trunks.
S: Two swimming trunks can not be added, otherwise the bag's maximal
   weight limit imposed by the airline would be exceeded. You have to
   care for enough space in the bag before further items can be chosen.
   This can be done by unpacking already packed items. If requested you
   get a listing of the already packed items.
U: One swimming trunk, does it fit?
S: The item swimming trunk can not be added, otherwise the bags maximal
   weight limit would be exceeded.
U: What do I have in my bag?
S: The following items were already selected: One macintosh, one jacket,
   ...
   [...]
```

Stage 3: Barrier “Climate Explanation"

```
S: The transmission of additional information about the holiday
   destination Waiuku is completed. Waiuku is located on the northern
   island of New Zealand in the Franklin district. [...] Do you wish
   this information to be repeated?
U: No.
S: The months November until February are the summer months in
   New Zealand. With an average temperature of 20 degree Celsius they are
   the warmest. In the winter months June until August average
   temperatures of only five to eight degree Celsius are reached. [...]
   For the time of your holidays the forecast gives temperatures about two
   to five degree and felt minus four to one degree Celsius. For the time
   of the forecast strong rain and hail are predicted.
```

Stage 4: Intervention `(Note: Half of the participants get an intervention like this.)`

```
S: Because of a broken data line the information about the destination
   could not be provided earlier. This may have changed your situation
   surprisingly. The selected items show that you expected different
   weather. If you knew the weather at the destination, you might have
   had selected different items. I am interested in your opinion about
   this.
U: Yes, I would have.
S: Did this situation evoke negative emotions? If yes, can you describe
   them?
U: [...]
S: I hope, your motivation to work on this task did not decrease too much.
   [...]
```

3.3 Questionnaires

To validate and extend the data collected during the experiment different psychological constructs are measured post hoc with the following questionnaires (german versions), too:

– Attributional-Style-Questionnaire (ASF-E [14]) measuring individual differences in attributional style,
– NEO-FFI [2] based on the Five Factor theory of personality,
– Inventory of Interpersonal Problems (IIP [10]) based on Sullivan's model of personality and focusing interpersonal relationships,
– Stress-coping questionnaire (SVF [11]) measuring different types of response to an unspecific selection of situations that impair, adversely affect, irritate or disturb the emotional stability or balance of the subject,
– TA-EG [12] measuring the technophily,
– Emotion Regulation Questionnaire (ERQ [7]) to ask for the two concepts of emotion regulation, the strategy of cognitive reappraisal and the suppression of emotional expression,
– BIS/BAS [1] based on Gray's theory of the Behavioral Inhibition System (BIS) and the Behavioral Approach System (BAS),
– AttrakDiff [9] measuring the subjective rating of usability and user experience.
– In addition to these psychometric instruments sociodemographic variables are collected like age, gender, educational level, experience with computers (e.g. years overall, hours per day/week) and in what context the subjects use the computer.

Based on the data from the experiments and the self-rating questionnaires different user types shall be differentiated.

3.4 Interviews

Furthermore a subgroup of participants undergoes an additional semi-structured interview of 60 to 160 Minutes after the experiment to determine their intentional stance on the computer system, their emotions during HCI and to identify further characteristics in their subjective experience of the HCI that support the differentiation of user-types.

At the beginning a so called narrative stimulus is given to evoke a narration about the subjective experiences during the WOZ-experiment: "*You have just done an experiment. Please try to recall this experiment. Please tell me, how you did during the experiment. Please tell me in detail, what you thought and experienced in the different situations!*". Afterwards, the interviewer tries to fill up gaps in the narration by so called immanent questions. Further exmanent questions focus on: occurred user emotions, intentional ascriptions towards the CS, the experience of the speech based interaction, the experience of the intervention (if given), the role of technique in autobiography and the general evaluation of the simulated CS. Methods of reconstructive qualitative social research that aim at the comprehension of the subjective experience of social reality and the related subjective meaning (here with respect to HCI) [5] will be used to analyse the interview material.

4 Evaluation

4.1 First Observations

As first inspections of the recordings show users very quickly adapt to restrictions that they experience. For example: wizards deliberately do not process conjunctions. If a user uses a conjunction in a selection (e.g. "*Three t-shirts and two jeans.*") then only the first conjunct is processed and the system gives resp. feedback ("*Three t-shirts have been added.*"). Most, but not all, users seem to quickly adapt to such implicit restrictions.

First unsystematic analyses of the interviews' material showed, that different feelings occur during the interaction with the system. Many subjects report annoyance, e.g., because the system did not warn about the weight limit in advance but always informed about its exceedance only after the subject spent time on trying to pack another item. Other subjects report uncertainty, e.g. when the system wants them to introduce themselves ("*I thought my god what shall I tell now*").

Furthermore different intentional ascriptions occur, e.g., one subject experienced the system as trustworthy in informing about the weather conditions ("*I trusted the computer when he said that it is that cold there, because of that I repacked my suitcase*"). Another ascription is malice, especially when the system gives additional information about the weather conditions at the holiday resort ("*I thought, well he wants to screw me, doesn't he*", "*now he wants to make fun of me*").

Metaphors, that many subjects use to explain how they experienced the system, seem to be very interesting to get insights into the user's subjective experience of the system, too. For example, one subject experienced the system as "*a puppet, that wants to speak with someone*" (in regard to the initial phase of the experiment).

Most of the subjects seem to try to adapt to the system, e.g., one subject reported that she had to think about what to answer when the system asked for events in the use of technical devices she experienced as positive and as negative ("*I had to pore over what I can tell now, what the machine is able to understand*").

4.2 Interaction Analysis of Transcripts

The subjects in our WOZ experiments only get told that they will interact in spoken language with a new type of system that shall be personalised and that therefore will ask some questions and pose some tasks. They do not get explicit instructions about the linguistic constructions and interaction patterns that are possible or not possible to use in the interaction.

How do people with differing technical background interact in such a situation?

One might speculate that there are a number of different possible reactions in such a situation. People could on the one hand actively try to find out the possibilities and limitations of the system by starting with – from their perspective – simple constructions and then gradually further trying more complicated ones. On the other hand people could simply rely on their intuitive assumptions (their internal picture based on whatever prior experiences or projections) about such possibilities and limitations. In the latter case no active exploration will be tried and behaviour – even if unnatural or inefficient – will be pursued as long as it proves to be successful. An example: One subject e.g. told in the post-hoc interview that in a situation where he did not know the term "*Konfektionsgrösse*" (engl. size of clothing) that he was asked about by the system he would have preferrred to do a Google search. "*But you could not ask the system for this!*", he continued, without having ever tried.

For the evaluation of the records and transcripts of the interaction there are many interesting questions. In the following we will name just a few.

Do users differ in their curiosity and there openness for experimentation? If so, how does this relate to results from the psychometric questionnaires?

Do users – although system initiative is dominant – at least try to take initiative by e.g. trying to ask questions? If so how do they react when the system is ignoring such attempts?

Do users e.g. try to interrupt lengthy system outputs (barge-in)?

Do users mirror the language of the system, e.g. on the lexical or syntax level? The system e.g. uses the general and somewhat uncolloquial term "*hinzufügen*" (engl. to add) in its feedback for selection operations. Do users mirror this usage?

How quickly do users adapt to observed limitations of the system? An example: are there any re-trials once a user experiences that from a conjunction only the first conjunct gets processed by the system?

How quickly do users employ efficient communication strategies like the use of ellipses in consecutive dialogue turns? Are there any subjects that stick to full blown sentences in a sequence of dialogue turns with a high potential for ellipses?

Do users employ synonyms (or hypernyms or other semantic equivalents) when they choose objects from the selection menus or – even more to be expected – when they unpack items from the suitcase? How do they react when the system refuses to understand a semantically correct unpacking command that does not use the term from the selection menu but a semantically equivalent one (e.g. "*remove three pairs of socks*" instead of "*remove three pairs of stockings*")?

Do users employ a kind of simplified language, e.g. without proper inflection, as they would use towards foreigners without sufficient knowledge of German?

These are some of the questions which will be focussed on in the first round of in-depth analyses.

5 Conclusion

At the time of writing ca. 60 of the planned 120 experiments of the current series are performed (including questionnaires and post hoc interviews). According to our schedule all 120 experiments will be completed in April 2011. The verbal interactions are transcribed in close relation to the resp. WOZ sessions.

Besides the aspects explained above the collected data will be used to explore the users' subjective experience during the interaction. These findings will also contribute to the description of user types in the interaction with companion systems on the one hand, and on the other hand to the development of adaptive dialogue strategies. Both shall contribute that the users' willingness to cooperate even in critical situations of the interaction is preserved.

On the basis of the reconstructed subjective experience of the interaction (especially the occurring emotions and ascriptions) dialogue strategies of the system should be identified, that are experienced as empathic or cooperation-hindering. This will contribute to develop dialogue strategies to support the development of positive ascriptions and to avoid the development of negative ones.

The current series of experiments will provide us with data and insights that stipulate future experiments. Based on the technical infrastructure and on the framework set up we will be able to explore systematically variants of the szenario. The overall issue will be, how subjects do experience the interaction with the (simulated) system when relevant system parameters are changed (e.g. quality of the TTS employed, allowed language fragment, varying NLU capabilities, talking style, initiative type, exhibited intelligent behaviour of the system).

In sum: We expect that the findings, insights and conclusions from our experiments will provide input for the ongoing efforts to develop the next generation of companion systems.

Acknowledgment

The presented study is performed in the framework of the Transregional Collaborative Research Centre SFB/TRR 62 "Companion-Technology for Cognitive Technical Systems" funded by the German Research Foundation (DFG). The responsibility for the content of this paper lies with the authors. Special thanks goes to the participants of the Workshop "Linguistische Analysen in der Mensch-Maschine-Interaktion" (Linguistic analyses in human-machine-interaction) in Magdeburg, Oct. 15–16, 2010.

References

1. Carver, C.S., White, T.L.: Behavioral Inhibition, Behavioral Activation, and Affective Responses to Impending Reward and Punishment: The BIS/BAS Scales. Journal of Personality and Social Psychology 67, 319–333 (1994)
2. Costa, P., McCrae, R.: NEO-PI-R Professional Manual. Revised NEO Personality Inventory (NEO-PI-R) and NEO Five Factor Inventory (NEO-FFI). Psychological Assessment Resources, Odessa (1992)
3. Dennett, D.C.: Intentional Systems. Journal of Philosophy 68, 87–106 (1971)
4. Dennett, D.C.: The Intentional Stance. The MIT Press, Cambridge (1987)
5. Flick, U.: An Introduction to Qualitative Research, 4th edn. Sage Publications Ltd., London (2009)
6. Frommer, J., Haase, M., Lange, J., Rösner, D., Friesen, R., Otto, M.: Project A3 'Prevention of Negative Dialogue Courses' Wizard of Oz Experiment Operator Manual, SFB-TRR-62 working paper (2010) (unpublished)
7. Gross, J.J., John, O.P.: Individual Differences in Two Emotion Regulation Processes: Implications for Affect, Relationships, and Well-Being. J. Pers. Soc. Psychol. 85, 348–362 (2003)
8. Halliday, M.: A Brief Sketch of Systemic Grammar. System and Function in Language, 3–6 (1976)
9. Hassenzahl, M., Burmester, M., Koller, F.: AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In: Ziegler, J., Szwillus, G. (eds.) Mensch & Computer 2003, pp. 187–196. B.G. Teubner, Stuttgart (2001)
10. Horowitz, L., Alden, L., Wiggins, J., Pincus, A.: Inventory of Interpersonal Problems Manual. The Psychological Corporation, Odessa (2000)
11. Jahnke, W., Erdmann, G., Kallus, K.: Stressverarbeitungsfragebogen mit SVF 120 und SVF 78, 3rd edn. Hogrefe, Göttingen (2002)
12. Karrer, K., Glaser, C., Clemens, C., Bruder, C.: Technikaffinität erfassen – der Fragebogen TA-EG. Der Mensch als Mittelpunkt technischer Systeme. 8. Berliner Werkstatt Mensch-Maschine-Systeme 29, 196–201 (2009), zMMS Spektrum, Reihe 22
13. Otto, M., Friesen, R., Rösner, D.: Message Oriented Middleware for Flexible Wizard of Oz Experiments in HCI. In: HCI International 2011. vol. this volume (2011)
14. Poppe, P., Stiensmeier-Pelster, J., Pelster, A.: Attributionsstilfragebogen für Erwachsene (ASF-E). Hogrefe, Göttingen (2005)
15. Rogers, C.: A Theory of Therapy, Personality and Interpersonal Relationships as Developed in the Client-centered Framework. In: Koch, S. (ed.) Psychology: A Study of a Science, Formulations of the Person and the Social Context, vol. 3, McGraw Hill, New York (1959)
16. Wilks, Y.: Close Engagements with Artificial Companions: Key Social, Psychological and Design issues. John Benjamins, Amsterdam (2010)

Multimodal Emotion Classification in Naturalistic User Behavior

Steffen Walter[1], Stefan Scherer[2], Martin Schels[2], Michael Glodek[2], David Hrabal[1], Miriam Schmidt[2], Ronald Böck[3], Kerstin Limbrecht[1], Harald C. Traue[1], and Friedhelm Schwenker[2]

[1] Medical Psychology, University of Ulm, Germany
[2] Institute of Neural Information Processing, University of Ulm, Germany
[3] Chair of Cognitive Systems, Otto von Guericke University Magdeburg, Germany
steffen.walter@uni-ulm.de

Abstract. The design of intelligent personalized interactive systems, having knowledge about the user's state, his desires, needs and wishes, currently poses a great challenge to computer scientists. In this study we propose an information fusion approach combining acoustic, and bio-physiological data, comprising multiple sensors, to classify emotional states. For this purpose a multimodal corpus has been created, where subjects undergo a controlled emotion eliciting experiment, passing several octants of the valence arousal dominance space. The temporal and decision level fusion of the multiple modalities outperforms the single modality classifiers and shows promising results.

1 Introduction

In the future, technological cognitive systems will more and more find their way into human's everyday life by supporting us with helpful information and mediating intentions in a social and technical environment. These systems need to be adaptive towards the individual user states such that the system is able to appropriately react to the user's needs and wishes. Requirements for this visionary goal are a robust and reliable automatic classification of emotional states. Since emotional user behavior is a complex, multimodal and dynamic process, it is important to consider a variety of channels such as prosody, mimics, gestures and bio-physiologic data [8,12]. For the analysis of naturalistic user behavior we therefore suspect multimodal approaches to outperform methods using a single modality with respect to classification performance [9].

In the presented study, we investigate a Wizard-of-Oz design aiming at inducing emotions according to the valance, arousal and dominance (PAD) model [15] in the subjects by simulating typical situations found in human computer interaction. Among the manifold possibilities to induce emotions we chose, e.g. delays, misunderstandings, ignoring commands, time pressure, but also positive feedbacks. For the automatic classification analysis we chose two octants of the PAD space, namely "positive valence, low arousal, high dominance" vs. "negative valence, high arousal, low dominance". The modalities of choice include

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 603–611, 2011.

respiration, heart rate, electromyography (M. corrugator), skin conductance, as well as audio based prosody, energy, voice quality and standard speech features, and manually annotated analysis of the facial expression due to the Facial Action Coding System (FACS) [5].

2 Experimental Design

To study the emotional behavior of humans in the different octants of the PAD space, a certain number of participants has been recorded in a Wizard-of-Oz (WOz) experiment. This method allows to simulate the technical system in a human-computer interaction by a psychologist in a separate room, which is connected via cameras and microphones. A group of 10 volunteers participated in the experiment. This group is divided equally in four subsets: younger female (mean of age = 27.6), older female (m = 46.4), younger male (m = 25.2) and older male (m = 56.2). In each gender group, the threshold of the age is 40 years. In Sect. 2.1 the intrinsic setting of the experiment will be illustrated, with which we expect to induce the demanded emotional states[1].

2.1 Course of the Experimental Sequences

The experiment is composed of two rounds: the first is divided into five so-called experimental sequences (ES). In each ES (ES-1_1-ES-5_1), the subject plays a game of Concentration (also known as Memory) of varying difficulty. The second round comprises six ES (ES-1_2-ES-6_2). The two rounds are conjoined by a center part, in which images of the International Affective Picture System (IAPS) [11] are shown to the subject. Figure 1 shows the schematic configuration of the course of the experiment.

Table 1 lists the characteristics of the different experimental sequences. The first column denotes the ES code, the second column the expected emotional states of the subject (e.g. "PAD +−+" represents positive valence, low arousal and high dominance). In the third column, the size of the Concentration matrix is shown, increasing from ES-1 to ES-6. The fourth column specifies the difficulty of the game. The last column lists the quality of the feedback given to the subject.

3 Feature Extraction

3.1 Audio Features and HMM Sequence Encoding

Prior to the audio processing it is necessary to perform a speech/non-speech detection to isolate the portions of the sequences in which the subject utters commands, indicating the position of the next memory card to be flipped in a two dimensional grid. The commands comprise the uttering of a letter indicating

[1] The study was carried out according to the ethical guidelines of Helsinki (ethic committee of the university of Ulm: C4 245/08-UBB/se).

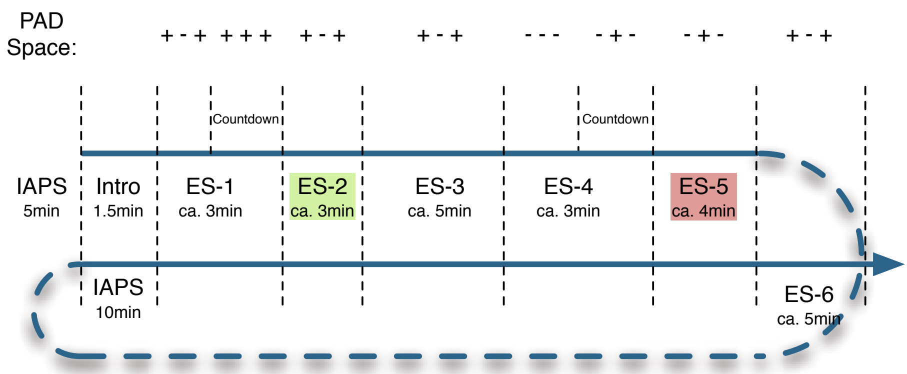


Fig. 1. Experimental design, including the expected position in the PAD space. After a five minute initialization with IAPS images, the experiment starts with a short introduction, followed by the first round of Concentration (ES-1,...,ES-5). The second round of Concentration contains six games (ES-1,...,ES-6) and starts after ten minutes presentation of IAPS images. The experiment is closed by standardized questionnaires. The expected octant of the PAD space is coded as plus and minus symbols in the order pleasure, arousal and dominance.

Table 1. Characteristics of the experimental sequences. For some sequences the arousal was increased (*) by introducing a countdown.

ES	Expected state	Matrix size	Difficulty	Feedback
ES-1 *	PAD +−+	4 × 4	low	slightly positive feedback (e.g. ”you have successfully solved the first task.”)
ES-2	PAD +−+	4 × 4	low	positive feedback (e.g. ”great!”)
ES-3	PAD +−+	4 × 5	low	neutral feedback (e.g. ”your performance is mediocre.”)
ES-4 *	PAD − − −	4 × 5	high	slightly negative feedback (e.g. ”your performance is declining”)
ES-5	PAD −+−	5 × 6	very high	negative feedback (e.g. ”you are doing very poorly.”)
ES-6	PAD +−+	5 × 6	low	positive feedback (e.g. ”your performance is improving.”)

the row of the board and a number indicating the column. Hence, the acoustic information is very limited and traditional features such as Mel frequency cepstral coefficients (MFCC) might not perform sufficiently well. However, since the commands are quite frequent and very similar in their structure and length, features such as the rate of speech, the length of the uttered letter or number, or the fundamental frequency could contain information on the subject state in the given task.

In order to investigate the distinction between positive and negative subject states, we computed seven diverse feature sets for the commands. Each of the sets is used to train an independent classifier. The decision for all the classifier is combined in a fusion step by simply multiplying the outputs of the respective classifier for each utterance. In the following we briefly introduce the feature sets:

1. The MFCC and ΔMFCC[2] features are extracted from 32ms sized windows with a sample rate of 100Hz. The features are design with respect to the biological perceptual variations in the human ear in order to capture the phonetically important characteristics of speech [3].
2. The biologically inspired Modulation spectrum (ModSpec) based features reflect the variations in the frequencies of the signal and are extracted from 200ms sized windows with a sample rate of 50Hz [10,6].
3. Perceptual linear predictive (PLP) features are an extension to the well known linear predictive coding features with respect to human perceptual capabilities comprising similar filtering as for MFCC and equal loudness curves [7,14]. The features are extracted from 32ms of speech and have a sample rate of 100Hz.
4. Statistics of the fundamental frequency (f_0) are calculated from each of the commands to form one feature vector [14]. The statistics comprise minimum, maximum, span, mean, and standard deviation as well as the same statistics on the differentiation of the f_0 signal.
5. Corresponding to the computation of the f_0, statistics of the energy of the signal are computed for each command.
6. Based on the voiced parts of the commands, we calculate statistics on the durations and pauses in between the latter and the number of the command using the f_0 signal. These features form the seventh set of audio based features, namely the speech rate and syllable duration features.

For the feature sets MFCC, ΔMFCC, ModSpec and PLP we calculate the distance matrices using a hidden Markov model encoding in order to compensate varying command lengths [17]. These encodings are then used as the actual features for the classification.

3.2 Facial Expressions Utilizing FACS Coding

The Facial Action Coding System (FACS) introduced in [5] systematically describes facial expressions. Each facial expression is decomposed by experts into so-called Action Units (AU), which encode contractions or relaxations of facial muscles (e.g. AU1: inner brow raiser, AU12: lip corner puller). These encodings could directly be useful as a high level feature input to the automatic classification.

3.3 Bio-physiological Feature Extraction

In this study we are provided with a variety of bio-physiological signals of the subject. From these signals sampled at 512 Hz, features were extracted for classification on the basis of a five second time window.

[2] The derivative of the Mel frequency cepstral coefficients.

One of the most prominent physiological features is the skin conductance level (scl), measuring the transpiration of a person. The gradient of this channel was calculated and thereupon the minimum, maximum and average gradients in the 5 second window are used for classification. A further intuitive feature for emotion recognition is the heart rate of a subject. In order to determine this characteristic, the blood volume pulse was recorded, which reflects the well known QRS structure of a heart beat. For every 5 second time window the minimum, maximum and the average rate of occurrences of R artifacts was calculated. For the EMG channels of the zygomaticus and the corrugator, we have first calculated the result of the moving average algorithm with window size 20 for each point of the segment to acquire a value for the tension of the muscle. Again, the minimum, maximum and average values of these signals were passed to classification.

4 Methods for Classification

Within this study two kinds of classifiers, namely Multilayer Perceptrons (MLP) and a Support Vector Machine (SVM) are used. This section gives a brief introduction to those classifiers. For a detailed description we recommend the reader to follow the citations.

5 Multilayer Perceptron

The multilayer perceptron (MLP), a universal function approximator, maps an input $\mathbf{x}$ to an output $\mathbf{y}$ processing the input via weighted ($\mathbf{w}$) connections through potentially multiple hidden layers [1]. In general, each perceptron calculates a weighted sum of all incoming variables plus an additional bias and then creates the output by mapping the value using a differentiable function called activation function. A simple perceptron without a hidden layer and a sigmoid activation function is therefore given by

$$y=\sigma(\sum_i x_i w_i + w_0). \tag{1}$$

An MLP is trained using error back-propagation and a gradient based weight adaptation after presenting labeled examples to the network and calculating an error function.

5.1 Probabilistic Outputs for Support Vector Machines

In recent years, SVM have found a broad acceptance in the Machine Learning Community. The success might have its origin in the intuitive generalization approach using a maximum margin between the classes and the capability to reformulate the SVM such that kernels can be used to transfer the data points implicitly into a higher dimensioned feature space. The SVM output in its original formulation renders crisp class decisions depending on which side of the margin the data point is located [19]. In many applications, especially in information fusion tasks, it can be advantageous to know the degree of belief that the output belongs to a certain class. Therefore, Platt [13] proposed an extension to SVM in which the distance of a point $\hat{\mathbf{x}}$ to the hyperplane is mapped onto a probability $p(\mathsf{t}|\mathbf{y} = y(\hat{\mathbf{x}}))$ using a sigmoid function where the distance to the hyperplane is given by $y(\hat{\mathbf{x}}) = \mathbf{w}^{\mathrm{T}}\hat{\mathbf{x}} + b$ and $\mathsf{t} \in \{-1, +1\}$. In order to fit the posterior the logistic sigmoid

$$p(\mathsf{t}|\mathsf{y} = y(\hat{\mathbf{x}}))=\frac{1}{1 + \exp(-ay(\hat{\mathbf{x}}) + d)} \tag{2}$$

having the parameters a and d is optimized using the cross-entropy error function.

5.2 Information Fusion

Combining information from multiple sources is a powerful means to stabilize classifiers but also to correct decisions of individual classifiers. In principle, there are two distinct ways of information fusion considering classification: firstly one can combine on a lower level of the general classification architecture, called *early* fusion , e.g. on a feature level. On the other hand, fusion is often considered on a decision level: the outputs of individual classifiers are combined in order to correct a more precise new decision. This approach is called *late* fusion or multiple classifier system (MCS) [9].

Generally information fusion is promising, when the combined sources show independence given the true class label, which is commonly called diversity in the MCS community [2]. Also, in case of an application using sequential data, e.g. physiological data or audio data, which are both used in this study, the temporal accumulation of individual decisions is advantageous. Thus, if the sequence of classifications have only a small bias to the correct class, an over-all correct classification can be accomplished [4]. Classifier fusion can be conducted using static combination rules as described in [18], but also using more complex trainable mappings [16].

6 Classification Experiments

Due to the heavily varying subject dependent bio-physiological data, we do not expect to be able to generalize over multiple subjects at once. Therefore, we conduct personalized experiments. As described above, we possess recordings of each individual undergoing the positive and negative experimental sequences twice (compare Section 2.1). Hence, in order to provide statistically salient results we conduct four cross-evaluation runs, i.e. training on round one and testing on the second, vice versa and folding the sequences of the two rounds.

For every audio feature, SVM with probabilistic outputs were trained and its results were combined on a decision basis using the averaging rule. These audio based results on the single utterance level led to an accuracy of 74.3 % with

Table 2. Confusion matrix of audio based utterance level late fusion

\	ES-2	ES-5
ES-2	0.477	0.523
ES-5	0.083	0.917

Table 3. Confusion matrix of bio-physiological five second window based classification using early fusion

\	ES-2	ES-5
ES-2	0.420	0.580
ES-5	0.474	0.526

Table 4. Confusion matrix of audio based sequence level fusion computed with temporal fusion

\	ES-2	ES-5
ES-2	0.556	0.444
ES-5	0.028	0.972

Table 5. Confusion matrix of bio-physiological sequence level classification computed with temporal fusion

\	ES-2	ES-5
ES-2	0.861	0.139
ES-5	0.417	0.583

Table 6. Confusion matrix after combining both modalities

\	ES-2	ES-5
ES-2	0.583	0.417
ES-5	0.028	0.972

an F_1 value of 0.600 for the positive class and 0.812 for the negative one over all the subjects and evaluation runs. The confusion matrix for the test data set is seen in Table 2. The bio-physiological data was combined using early fusion, i.e. concatenation of the features, and classified using a MLP. The results on the five second window classification led to an accuracy of 49.2 % and an F_1 value of 0.348 for the positive class and 0.584 for the negative class over all the subject and evaluation runs. The confusion matrix for this experiment is found in Table 3.

In the following temporal fusion is conducted using an average over the particular sequence. The results of this temporal fusion are shown in the confusion matrices seen in Table 4 and 5. For the audio classification, this procedure resulted in an over all accuracy of 76.4 % and an F_1 value of 0.702 for the positive class and 0.805 for the negative class. In case of the bio-physiological data, we achieved an accuracy of 72.2 % and an F_1 value of 0.756 for the positive sequences and 0.677 for the negative class.

These temporally combined decisions for both modalities audio and bio-physiology are further combined in a decision fusion step and yield the following results: an overall accuracy of 77.8% is achieved with an F_1 score of 0.712 for the positive class and 0.805 for the negative one.

Additionally, a certified FACS-coder observed all changes in facial musculature as described in the FACS manual. Conducted Wilcoxon rank sum tests showed statistically significant differences of the distribution of the occurring action units AU1 ($p = 0.042$), AU10 ($p = 0.041$), AU14 ($p = 0.017$), and AU28 ($p = 0.020$) when comparing ES-2 and ES-5. Moreover, the observed mean values were significantly lower in ES-2 than in ES-5.

7 Summary and Discussion

In the presented study information fusion techniques are applied in order to improve results for multimodal emotion classification. We conducted an utterance based classification using seven distinct audio feature sets yielding a high accuracy of 74.3 %. On the other hand using bio-physiological data only 49.2 % of the five second long clips were correctly classified, suggesting that emotion recognition in the utilized channels requires more information than the one that is present in these short windows.

In a further step a temporal fusion of both modalities was conducted, improving both results significantly. Especially the bio-physiological approach improved to an accuracy of 72.2 %.

Observing the confusion characteristics seen in Table 4 and 5, reveals opposing error behavior. Hence, a decision fusion of both could yield much improved results. The decision fusion, however, only slightly improved the overall result, which indicates that more effort needs to be put into this particular step.

The results considering the facial expression analysis have not yet incorporated into the information fusion framework. The presented tests nonetheless reveal a significant differences between the considered classes, which makes the video modality a promising source for classification in this application.

Acknowledgment

This research was supported in part by grants from the Transregional Collaborative Research Centre SFB/TRR 62 "Companion-Technology for Cognitive Technical Systems" funded by the German Research Foundation (DFG). The work of Martin Schels is supported by a scholarship of the Carl Zeiss Foundation. Miriam Schmidt is supported by a scholarship of the graduate school Mathematical Analysis of Evolution, Information and Complexity of the University of Ulm.

References

1. Bishop, C.M.: Pattern Recognition and Machine Learning (Information Science and Statistics), 1st edn. Springer, Heidelberg (2006)
2. Brown, G., Kuncheva, L.I.: "Good" and "Bad" diversity in majority vote ensembles. In: El Gayar, N., Kittler, J., Roli, F. (eds.) MCS 2010. LNCS, vol. 5997, pp. 124–133. Springer, Heidelberg (2010)
3. Davis, S., Mermelstein, P.: Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. IEEE Transactions on Acoustics, Speech and Signal Processing 28(4), 357–366 (1980)
4. Dietrich, C., Schwenker, F., Palm, G.: Classification of time series utilizing temporal and decision fusion. In: Kittler, J., Roli, F. (eds.) MCS 2001. LNCS, vol. 2096, pp. 378–387. Springer, Heidelberg (2001)
5. Ekman, P., Friesen, W.V.: Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press, Palo Alto (1978)

6. Hermansky, H.: The modulation spectrum in automatic recognition of speech. In: Proceedings of IEEE Workshop on Automatic Speech Recognition and Understanding, pp. 140–147. IEEE, Los Alamitos (1997)
7. Hermansky, H., Hanson, B., Wakita, H.: Perceptually based linear predictive analysis of speech. In: Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 1985), vol. 10, pp. 509–512 (1985)
8. Kim, J., André, E.: Emotion recognition based on physiological changes in music listening. IEEE Trans. Pattern Anal. Mach. Intell. 30, 2067–2083 (2008), http://portal.acm.org/citation.cfm?id=1477073.1477535
9. Kuncheva, L.I.: Combining Pattern Classifiers: Methods and Algorithms. Wiley Interscience, Hoboken (2004)
10. Maganti, H.K., Scherer, S., Palm, G.: A novel feature for emotion recognition in voice based applications. In: Paiva, A.C.R., Prada, R., Picard, R.W. (eds.) ACII 2007. LNCS, vol. 4738, pp. 710–711. Springer, Heidelberg (2007)
11. Peter, J., Lang, M.M.B., Cuthbert, B.N.: International affective picture system (iaps): Affective ratings of pictures and instruction manual. Tech. rep., NIMH Center for the Study of Emotion & Attention, University of Florida (2008)
12. Picard, R.W.: Affective Computing. MIT Press, Cambridge (2000)
13. Platt, J.: Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. Advances in Large Margin Classifiers, 61–74 (1999)
14. Rabiner, L.R., Schafer, R.W.: Digital processing of speech signals. Prentice-Hall Signal Processing Series. Prentice-Hall, Englewood Cliffs (1978)
15. Russell, J.: A circumplex model of affect. Journal of Personality and Social Psychology 39, 1161–1178 (1980)
16. Schwenker, F., Dietrich, C., Thiel, C., Palm, G.: Learning of decision fusion mappings for pattern recognition. International Journal on Artificial Intelligence and Machine Learning (AIML) 6, 17–21 (2006)
17. Smyth, P.: Clustering sequences with hidden markov models. Advances in Neural Information Processing Systems 9, 648–654 (1997)
18. Tax, D.M.J., van Breukelen, M., Duin, R.P.W., Kittler, J.: Combining multiple classifiers by averaging or by multiplying. Pattern Recognition 33(9), 1475–1485 (2000)
19. Vapnik, V.N.: The nature of statistical learning theory. Springer-Verlag New York, Inc., New York (1995)

Author Index